DOCUMENT RESUME

ED 076 955 CS 000 537

AUTHOR Salinas, Mary Ellen Swanson
TITLE 1967-1968 (Year Four) Findings: A Comparative Study of the Effects of Oral-Aural Teaching Techniques on Pupils' Gain in Reading, Language and Work Study Skills in Grades Three and Four.
PUB DATE 72
NOTE 251p.; Ph.D. Dissertation, University of Texas

EDRS PRICE MF-$0.65 HC-$9.87
DESCRIPTORS *Disadvantaged Youth; Elementary Grades; Grade 3; Grade 4; *Language Instruction; Longitudinal Studies; Mexican Americans; *Oral Communication; Reading; Reading Achievement; *Reading Research; *Spanish Speaking; Teaching Techniques
IDENTIFIERS *San Antonio Language Research Project

ABSTRACT

This study attempted to evaluate the effects on pupil performance of three and four years of intensive oral language instruction in English or Spanish. The subjects were disadvantaged third and fourth grade Mexican-American children who had been members of experimental classes since first grade in the San Antonio Language Research Project. The attempt was made to include in the data analysis scores only from children who had been in the same type of experimental treatment for all their years in school. The Iowa Tests of Basic Skills were used to measure improved language abilities in reading. The nonverbal battery of the Lorge-Thorndike Intelligence Tests was used to estimate differences in learning capacity. The most consistent set of findings related to the ineffectiveness of the language-cognition Spanish treatment. On almost no measure did the language-cognition Spanish groups receive scores higher than the other experimental or control groups. The scores of the children in all the treatments, experimental and control, at both grade levels were very low. (WR)

FILMED FROM BEST AVAILABLE COPY

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 076955

1967-1968 (Year Four) Findings, San Antonio Language Research Project

A Comparative Study of

The Effects of Oral-Aural Teaching Techniques on Pupils' Gain in Reading, Language and Work Study Skills in Grades Three and Four

by Mary Ellen Swanson Salinas

THE UNIVERSITY OF TEXAS AT AUSTIN · AUSTIN, TEXAS 78712

ED 076955

1967-68 (YEAR FOUR) FINDINGS: A COMPARATIVE STUDY OF THE EFFECTS OF ORAL-AURAL TEACHING TECHNIQUES ON PUPILS' GAIN IN READING, LANGUAGE AND WORK STUDY SKILLS IN GRADES THREE AND FOUR

APPROVED BY SUPERVISORY COMMITTEE:

Thomas J. Horn

Roland A. Arnold

Dennis R. Laurila

Daniel C. Morgan, Jr.

Copyright

by

Mary Ellen Swanson Salinas

1972

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Mary Ellen
Swanson Salinas

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

1967-68 (YEAR FOUR) FINDINGS: A COMPARATIVE STUDY OF THE EFFECTS OF ORAL-AURAL TEACHING TECHNIQUES ON PUPILS' GAIN IN READING, LANGUAGE AND WORK STUDY SKILLS IN GRADES THREE AND FOUR

by

MARY ELLEN SWANSON SALINAS, B.A., M.A.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August, 1972

PREFACE

This study by Mary Ellen Swanson Salinas was originally intended as primarily an analysis of the San Antonio Language Research Project data for its fourth year (1967-68). However, due to the fact that data analysis was delayed until the 1971-72 year, the scope of the report was expanded to include a review and evaluation of the entire project on a longitudinal basis. Mrs. Salinas is to be commended for her objective, straightforward, frank and honest, yet sympathetic, analyses of the trials, tribulations, successes (admittedly few) and failures of a relatively early under-financed attempt at longitudinal research in educational innovations for children of poverty.

In the words of Theodore Sizer (Phi Delta Kappan, June, 1972, p. 635), former dean of the Harvard Graduate School of Education and headmaster to be at Phillips Academy, Andover,

> "To be honest with ourselves in education is heady medicine, requiring not only courage but self-confidence. We must be sure that we can, when faced with fresh evidence or new conditions, shift our manner of work--even our beliefs--fundamentally."

Thomas D. Horn

The University of Texas at Austin
August, 1972

ACKNOWLEDGMENTS

I wish to express appreciation and gratitude to the members of my supervising committee, all of whom went considerably beyond the call of duty to help me in my graduate work and with this study: Thomas D. Horn, Muriel R. Saville and Dan C. Morgan, Jr., of The University of Texas at Austin and Richard D. Arnold of Purdue University.

I also wish to thank Kathy Silva, Johanna Starcich, Lee Edmonston, Claire Weinstein, Ellen Snow, Sue McClellan, Jean Goodloe, José Nicandro Garza, Mary Alice Ramírez, Liz Davis, Earl Jennings, Bettie Carrington, Edmund Emmer, and Suanne Cardwell for their help. Dan Blaine and Hugh Poynor rendered special assistance with regard to the data analysis and interpretation. The writers of other studies on the San Antonio Language Research Project and personnel of the San Antonio Independent School District supplied additional help. I am especially thankful for the advice and support provided by Walter F. Stenning.

Finally, I am indebted to my parents, Dr. and Mrs. L. W. Swanson, and to my husband, Fernando, for their encouragement and extreme patience.

M.E.S.S.

The University of Texas at Austin
August, 1972

TABLE OF CONTENTS

CHAPTER		PAGE
I	INTRODUCTION	1
	Importance of the Study	8
	Statement of the Problem	12
	Instrumentation	12
	Sample Description	13
	Hypotheses	15
	Design and Method of Analysis	17
	Organization of Rest of the Study	17
II	REVIEW OF PROJECT LITERATURE	18
	Historical Perspective	18
	Year One (1964-65) Findings	28
	Year Two (1965-66) Findings	36
	Year Three (1966-67) Findings	45
	Year Five (1968-69) Findings	49
	Related Studies	59
	Studies of Oral Language	59
	Ott	59
	Jameson	75
	Peña	82
	Studies on Reading	94
	Arnold	94
	Fowler	96
	Pauck	102

CHAPTER PAGE

Melton..... 108

Cook..... 110

Socioeconomic Factors..... 113

MacMillan..... 113

McDowell..... 117

Additional Studies..... 124

Summary..... 125

III DESCRIPTION OF PROCEDURES AND RESULTS OF DATA ANALYSES..... 135

Description of Procedures..... 136

Population..... 136

Sampling Procedure 136

Instrumentation..... 141

The Lorge-Thorndike Intelligence Tests... 146

Iowa Tests of Basic Skills..... 148

Testing Procedures..... 153

Data Analysis Procedures..... 154

Results of Data Analyses..... 154

Descriptive Statistics..... 154

Comparing Treatment Groups by Subtest.... 157

Comparing Subtest Scores of Treatment Groups..... 158

Inferential Statistics..... 163

Fourth Grade..... 163

Third Grade..... 177

CHAPTER PAGE

IV SUMMARY, LIMITATIONS, CONCLUSIONS, DISCUSSION AND RECOMMENDATIONS..... 209

Summary..... 209

Findings..... 212

Fourth Grade..... 212

Third Grade..... 213

Long-Term Patterns..... 217

Sample I..... 217

Sample II..... 221

Limitations..... 222

Instrumentation..... 222

Sample Size..... 223

Lack of Monitoring..... 223

Nature of San Antonio Curriculum..... 224

Assessment of Oral Language..... 225

Assignment to Treatment Groups..... 226

Retention Policy..... 226

Teacher Variables..... 227

Conclusions..... 228

Discussion..... 230

Spanish Treatment Findings..... 230

English Treatment Findings..... 231

San Antonio Curriculum Findings..... 232

NOA-LCE Treatment Findings..... 233

Influence of IQ..... 234

Overall Results..... 236

CHAPTER PAGE

Worth of the Project..... 237

Recommendations..... 238

APPENDIX A..... 241

APPENDIX B..... 243

BIBLIOGRAPHY..... 244

LIST OF TABLES

TABLE		PAGE
1	Number of Children Remaining in Original Treatment Groups in Year Five (1968-69) Compared to Number for Whom Complete Data Were Collected in First Grade, San Antonio Language Research Project.....	50
2	Summary of Order of Means for Phonology, Intonation, Fluency, and Total Language, Year Five (1968-69), San Antonio Language Research Project.....	55
3	Correlation Coefficients Computed on All Students in the Sample (N=161) for Phonology, Intonation, Fluency, and Total Language, Year Five (1968-69), San Antonio Language Research Project.....	57
4	F-Ratios and Probability Levels for Three Comparisons of Ott-Jameson Test and Linguistic Capacity Index Variables for First Graders in OAE (Oral-Aural English) and NOA (Non-Oral-Aural) Treatment Groups, Year Two (1965-66), San Antonio.....	71
5	Number of Children in Experimental Classes and in Data Analyses Year One (1964-65), Year Two (1965-66) and Year Three (1966-67), San Antonio Language Research Project.....	138
6	Number of Children in Research Sample Year Four (1967-68), San Antonio.....	140
7	Number of Teachers and Schools in Research Sample, Year Four (1967-68), San Antonio.....	142
8	Means, Standard Deviations, F-Ratios and Probability Levels for Spring Administration of Iowa Tests of Basic Skills for Grade Four in Year Four (1967-68), San Antonio.....	155
9	Means, Standard Deviations, F-Ratios and Probability Levels for Spring Administration of Iowa Tests of Basic Skills for Grade Three in Year Four (1967-68), San Antonio.....	156

TABLE		PAGE
10	Percentile Ranks of Grade-Equivalent Averages on Iowa Tests of Basic Skills for Grade Four in Year Four (1967-68), San Antonio.....	160
11	Percentile Ranks of Grade-Equivalent Averages on Iowa Tests of Basic Skills for Grade Three in Year Four (1967-68), San Antonio.....	161
12	Comparison of Fourth and Third Grade Mean Scores of Each Treatment Group on Iowa Tests of Basic Skills for Year Four (1967-68), San Antonio.....	206
13	Summary of Results of Analysis of Variance and Analysis of Covariance, San Antonio Language Research Project, Grade Four, Year Four (1967-68).....	214
14	Summary of Results of Analysis of Variance and Analysis of Covariance, San Antonio Language Research Project, Grade Three, Year Four (1967-68).....	216
15	Summary of Findings Related to Sample I (children who entered first grade in 1964) in the San Antonio Language Research Project at the End of Year Four (1967-68).....	218
16	Summary of Findings Related to Sample II (children who entered first grade in 1965) in the San Antonio Language Research Project at the End of Year Four (1967-68).....	219

LIST OF FIGURES

FIGURE PAGE

1 Least-Squares Regression Lines for Each of the Three Groups (Grade Four, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Vocabulary Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 164

2 Least-Squares Regression Lines for Each of the Three Groups (Grade Four, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Spelling Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 167

3 Least-Squares Regression Lines for Each of the Three Groups (Grade Four, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Language Skills--Usage Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 169

4 Least-Squares Regression Lines for Each of the Three Groups (Grade Four, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Language Skills--Total Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 171

5 Least-Squares Regression Lines for Each of the Three Groups (Grade Four, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Composite Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 176

FIGURE PAGE

6 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Vocabulary Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 180

7 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Reading Comprehension Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 182

8 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Language Skills--Spelling Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 183

9 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Language Skills--Capitalization Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 185

10 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Language Skills--Punctuation Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 187

FIGURE PAGE

11 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Language Skills--Usage Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 189

12 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Language Skills--Total Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 191

13 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Work-Study Skills--Map Reading Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 192

14 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Work-Study Skills--Reading Graphs and Tables Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 194

15 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Work-Study Skills--Knowledge and Use of Reference Materials Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 196

FIGURE PAGE

16 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Work-Study Skills--Total Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 198

17 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Arithmetic Skills--Total Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 201

18 Least-Squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Composite Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ..... 202

CHAPTER I

INTRODUCTION

A sample survey made in November, 1969, by the U.S. Bureau of the Census found that 9,230,000 persons--about five percent of the approximately 200,000,000 people in the United States--reported that they were of Spanish origin,[1] e.g., Mexican, Cuban, Puerto Rican or from other Spanish-speaking countries. Comparisons regarding income and employment included the findings that:

> Spanish origin family income was on the average $5,600, about 70 percent of that of other families.
>
> Spanish origin persons had an unemployment rate of 6.0 percent, about 1.7 times the unemployment rate of all other persons in the labor force.
>
> Spanish origin workers were less likely to hold white-collar jobs than the remaining employed populations of the United States.[2]

Sixty-one percent of the people of Spanish origin lived in the Southwest, i.e., Texas, California, Colorado, New Mexico and Arizona.[3] Five million, seventy-three

---

[1]U.S. Bureau of the Census, Current Population Reports, Series P-20, No. 213, "Persons of Spanish Origin in the United States: November 1969," U.S. Government Printing Office, Washington, D.C., 1971, pp. 1-2. (Hereafter referred to as "Persons of Spanish Origin.")

[2]Ibid., p. 2.

[3]Ibid.

thousand or fifty-five percent of the persons of Spanish origin identified themselves as being of Mexican origin. Eighty-six percent of these people lived in the Southwest.[1] There was no breakdown by state, but another study estimated that eighty-five percent of the Mexican-Americans who live in the Southwest, live in California and Texas.[2]

Further evidence of population concentration is the fact that thirty-three and one-half percent of the almost three and one-half million persons with Spanish surnames lived in fourteen cities in the Southwest, as of 1960.[3] Data on the Spanish population in the 1950 census and 1960 census were based on the less inclusive measure of "Spanish surname." Many persons of Spanish origin do not have Spanish surnames. Within Texas, persons with Spanish surnames are concentrated in the southern part of the state so that they are the majority ethnic group in cities such as Brownsville and Laredo.[4]

---

[1] Ibid., p. 4.

[2] Walter Fogel, "Mexican-Americans in Southwest Labor Markets," Mexican-American Study Project Advance Report 10, University of California, Los Angeles, 1967, p. 3.

[3] Robert W. MacMillan, "A Study of the Effect of Socioeconomic Factors on the School Achievement of Spanish-Speaking School Beginners," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1966, pp. 35 and 38.

[4] Herschel T. Manuel, Spanish Speaking Children of the Southwest, The University of Texas Press, Austin, 1965, p. 29.

Mexican-Americans have been characterized historically as a unique group because of their high degree of homogeneity. Being Mexican-American has been almost synonymous with being poor. Saunders reported in 1949 that:

> Everywhere . . . there is poverty. Not all Spanish-speaking are poor, but in general more of them are poor than is true for any other group. . . . If one were to attempt to characterize the conditions of the Spanish-speaking Texans, he would be forced to say that, in general, and for nearly any index of socio-economic status that might be devised, the Spanish-speaking people are found to occupy a less desirable position than the Anglos or the population as a whole.[1]

Conclusions by Bogue about Mexican-Americans in the 1950's revealed no substantial change from Saunders' observations of the 1940's.

> As a group, immigrants from Mexico have a very low educational attainment. Although the educational attainment of the second generation is somewhat higher than most of the first generation, it is still far below that of the native stock or of any other immigrant group. Mexicans are employed largely as unskilled laborers, and there is not a great deal of difference between first and second generations in this respect. The Mexican is the most poorly educated member of the nation's population, with an educational level lower than that of the Puerto Rican. This is the only ethnic group for which a comparison

---

[1]Lyle Saunders, Address given to the National Convention, League of United Latin American Citizens, San Antonio, Texas, June 11, 1949, pp. 7-8 (Mimeographed), quoted by Carey McWilliams, Brothers Under the Skin, Little, Brown and Company, Boston, 1951, pp. 131-2.

> of the characteristics of the first and second generation fails to show a substantial inter-generational rise in socio-economic status.[1]

Data from the 1960's show continued lack of progress. In November, 1969, 6.2 percent of the persons of Mexican origin who were 16 years of age or older and were in the labor force were unemployed. At that time, only 3.6 percent of all people in the labor force 16 years old and over were unemployed.[2]

In a classification of occupations that included farm workers, service workers, blue-collar workers and white-collar workers, only 18.5 percent of Mexican males were white-collar workers in November, 1969. This compared to figures of 25.3 percent for all males of Spanish origin and 41.4 percent for men of all other origins. Figures for female workers were as follows: 37.7 percent of those of Mexican origin, 40.9 percent of those of Spanish origin, and 60.7 percent of all other origins were white-collar workers.[3]

The median income for households with a head of Spanish origin in November, 1969, was $5,641--compared to

---

[1]Donald J. Bogue, The Population of the United States, The Free Press, Glencoe, Ill., 1959, p. 372.

[2]"Persons of Spanish Origin," p. 34.

[3]Ibid., p. 29.

$8,011 for all other origins and $5,488 for heads of Mexican origin.[1]

Clearly, the socioeconomic condition of the Mexican-American has not improved spectacularly in the last 30 years. Additional data can be found in Appendix A which makes the situation of the Mexican-Americans in the Southwest, Texas and San Antonio even more graphic.

Judging from what is known about the difficulties that economically disadvantaged children have in school, one would expect Mexican-Americans to have attained only minimal levels of educational achievement.[2] Data confirm this expectation. Of persons aged 20-49 in Texas, one-third of the Anglos have not graduated from high school, while over three-fourths of the Mexican-Americans have not. In seventh through twelfth grades the dropout rate of Mexican-Americans is almost twice that of Anglos, and Mexican-Americans drop out sooner.[3]

In addition to the problem of poverty that the typical Mexican-American child brings to school, he also

---

[1] Ibid., p. 34.

[2] Paul A. Witty (ed.), "The Educationally Retarded and Disadvantaged," Sixty-Sixth Yearbook for the National Society for the Study of Education, Part I, University of Chicago Press, Chicago, 1967.

[3] Governor's Committee on Public School Education, The Challenge and the Chance, Governor's Committee on Public School Education, Austin, Texas, 1963 (non-paged).

is likely to be deficient in English language skills. A 1957 report of the Texas Education Agency indicated that at that time, the average child in Texas with a Spanish surname was spending three years in first grade and was dropping out of school before reaching the fifth grade. This educational deficiency was correlated by Andersson with inability in English.[1]

Similarly, a survey involving 123 Texas school districts in 1962 found that:

> The failure rate of Spanish-speaking children without any preschool instruction in the first grade was 82 percent. A major reason for this failure was reported to be their inability to comprehend and use English.[2]

In 1964 in a predominantly Mexican-American area of San Antonio, Texas, 98 percent of the children entering first grade spoke no English.[3]

---

[1]Theodore Andersson, "Bilingual Education: The American Experience," Paper presented at the Ontario Institute for Studies in Education Conference on Bilingual Education, Toronto, March 13, 1971, p. 11 (Mimeographed), citing Texas Education Agency, Division of Research, "Reports of Pupils in Texas Public Schools Having Spanish Surnames, 1955-56," Texas Education Agency, Austin, 1957.

[2]Elaine D. Fowler, "An Evaluation of the Brengelman-Manning Linguistic Capacity Index as a Predictor of Reading Achievement of Spanish-Speaking First Grade Children," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1969, p. 1, citing Texas Education Agency, A Report on the Preschool Instructional Program for Non-English Speaking Children, Texas Education Agency, Austin, 1962.

[3]Thomas D. Horn, "A Study of the Effects of Intensive Oral-Aural English Language Instruction, Oral-Aural Spanish

Despite the recent proliferation of so-called bilingual education programs, the "problem" of children coming to school without well-developed English language skills is not likely to disappear in the United States in the near future.

As mentioned earlier, according to a sample survey made in November, 1969, there were 9,230,000 persons in the United States who identified themselves as being of Spanish origin. Of these, 6,700,000 reported Spanish as their mother tongue. Spanish was the current home language of 4,622,000 persons.[1] Of all persons of Mexican origin, 47.3 percent (or 2,401,000 persons) reported that Spanish is the language usually spoken in the home.[2]

Given the socio-economic history and present condition of many Mexican-Americans, combined with the census data regarding language currently spoken in the home, it is safe to assume that for some time to come many Mexican-American children in the Southwest will be doubly penalized in school--because of their socio-economic status and because of their English linguistic abilities.

Language Instruction and Non-Oral-Aural Instruction on Reading Readiness in Grade One," The University of Texas, Austin, 1966, p. 14.

[1] "Persons of Spanish Origin," p. 1

[2] Ibid., p. 14.

IMPORTANCE OF THE STUDY

The west side of San Antonio, Texas, is densely populated by disadvantaged Mexican-Americans. In an effort to precipitate change in the schools and to develop and demonstrate specific pedagogical techniques, the San Antonio Language Research Project was inaugurated in the fall of 1964 in ten west-side San Antonio schools. The Project was directed by Thomas D. Horn and had the cooperation of the San Antonio Independent School District.

Since the assumptions were that oral and listening English language skill levels of Mexican-American children were insufficient for successful academic achievement, the educational techniques used in the experimental treatments were various methods of oral language instruction. One treatment consisted of intensive oral-aural language instruction in English; another, intensive oral-aural language instruction in Spanish; and a third, unstructured language activities in English. All three treatments used science materials as content.

The original stated purpose and research focus of the Project in Year One (1964-65) was to "compare the effectiveness of three methods of developing reading readiness in Spanish-speaking boys and girls in the first grade."[1] While the efforts to improve oral language

---

[1]Horn, op. cit., pp. 3-4.

continued into successive years, the goals changed from reading readiness to reading proficiency and achievement, to academic achievement in general. The San Antonio Language Research Project was in effect exploring two major questions:

(1) Are structured audio-lingual techniques more effective than more informal language activities for improving oral English and Spanish?

(2) Will improvement in oral language be reflected in improved achievement in reading and other academic subjects?

Several doctoral dissertations, articles, and research reports concerning the Project have been completed. Four of these are viewed as reporting the results of the experimental treatments at the end of successive school years for children entering first grade in Years One (1964-65) and Two (1965-66).[1] Taylor, Year Five (1968-69), attempted to answer the first major question mentioned above while the authors of the other three yearly reports (Year One--1964-65, Year Two--1965-66, and

---

[1] Horn, op. cit.; Richard D. Arnold, "1965-66 (Year Two) Findings, San Antonio Language Research Project, Thomas D. Horn, Director," The University of Texas, Austin, 1968; Lester N. Knight, "A Comparison of the Effectiveness of Intensive Oral-Aural English Instruction, Intensive Oral-Aural Spanish Instruction, and Non-Oral-Aural Instruction on the Reading Achievement of Spanish-Speaking Second and Third-Grade Pupils," Unpublished Doctoral Dissertation,

Year Three--1966-67) assumed that language had improved and tried to provide answers to the second question.

If one is willing to assume that the tests used to measure reading and reading readiness were equally valid for all treatment groups, then the following conclusions can be made regarding the results during the first three years of the Project: after Year One (1964-65) there were no significant differences between the treatment groups with regard to reading readiness. The findings in Years Two (1965-66) and Three (1966-67) were less clear and somewhat contradictory. In general, however, the children who entered first grade in Year One (1964-65) were superior in reading achievement if they were in the treatment receiving unstructured (non-oral-aural) language activities.[1] The children who entered first grade in Year Two (1965-66) were superior if they received intensive oral-aural language instruction in English.[2] This could be due to benefits accruing from teachers' previous experience with the treatment.

The University of Texas, Austin, 1969; and Thomasine H. Taylor, "A Comparative Study of the Effects of Oral-Aural Language Training on Gains in English Language for Fourth and Fifth Grade Disadvantaged Mexican-American Children," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1969.

[1]Arnold, op. cit., p. 69; Arnold, op. cit., p. 215.

[2]Arnold, op. cit., p. 70; Knight, op. cit., p. 215.

In Year Four (1967-68) the informal language activities of the non-oral-aural treatment were dropped. Then two main treatment groups were left at each grade: students who had received four or three years of intensive oral language instruction in English or in Spanish. Due to lack of adequate instruments to measure growth in oral language, no empirical evidence on oral language development was collected. There was no alternative but to assume that structured audio-lingual techniques were more effective than typical language activities for improving oral language. The task became one of trying to answer the second major question: Will improvement in oral language be reflected in improved achievement in reading and other academic subjects?

By the end of Year Four (1967-68) three additional questions became of special interest. First, would the patterns of superiority of the non-oral-aural treatment for the children who entered first grade in Year One (1964-65) and of the intensive English treatment for the children who entered first grade in Year Two (1965-66) continue into Year Four (1967-68)? Second, would analysis of test scores after three or four years reveal the superiority of one treatment more clearly and dramatically than in previous years? Transfer of learning from oral language to reading may take place only after considerable training. Third, would the oral language training

have any effect in academic areas? Up to this point, no one had investigated the effectiveness of the experimental treatments on school subjects other than reading.

STATEMENT OF THE PROBLEM

The problem in this study was to evaluate the effectiveness of experimental oral language programs at the end of three and four years of instruction. This was done by comparing the levels of intellectual skills and abilities of pupils in experimental third and fourth grade classes who had received intensive oral language instruction in English or Spanish and of pupils in control classes in third and fourth grade who had received no intensive oral language instruction.

INSTRUMENTATION

The tests used to measure the level of intellectual skills and abilities were the Iowa Tests of Basic Skills, administered in the spring of Year Four (1967-68).[1] The Iowa Tests of Basic Skills measure skills in five major areas: vocabulary, reading, language, work-study and arithmetic. Four subtests measure skills in the language area: spelling, capitalization, punctuation and usage. The work-study area is divided into three sections: map

---

[1]E.F. Lindquist and A.N. Hieronymus, Houghton Mifflin Co., Boston, 1964.

reading, reading graphs and tables, and knowledge and use of reference materials. In arithmetic the two subtests are: arithmetic concepts and arithmetic problem solving. Scores on The Lorge-Thorndike Intelligence Tests, administered in the fall of third grade, were used to estimate differences in learning capacity.[1]

SAMPLE DESCRIPTION

The pupils tested in this study were in the Project schools during their entire school careers. It was established during Year One (1964-65) that, in almost every instance, these children were members of families of low socio-economic status and low educational attainment.[2] Their control of English at the beginning of first grade was informally assessed by several observers as being minimal or non-existent.[3]

During Year Four (1967-68) approximately 550 pupils in fourth grade and 675 pupils in third grade were members of experimental classes. Unfortunately, not all of these pupils had always been in the same type of experimental treatment, or in any experimental treatment. Moreover, a few of the children had failed one or more grades. When

---

[1]Irving Lorge and Robert L. Thorndike, Houghton Mifflin Co., Boston, 1957.

[2]MacMillan, op. cit., p. 134.

[3]Horn, op. cit., p. 14.

data for such children were eliminated, few pupils were left. Incomplete data eliminated still more pupils. Therefore, it was decided to include in the study data for all children identified as having received the same treatment for their three or four years in school. Random sampling from the experimental classes was neither necessary nor advisable.

The figures below show the number of children who were included in this study from the intensive English and intensive Spanish classes:

	Fourth Grade	Third Grade
LCE	29	109
LCS	60	40

Henceforward, the intensive English treatment will be referred to as the LCE treatment and the intensive Spanish treatment will be referred to as the LCS treatment.

At the third grade level, one additional group was studied. These pupils had been in non-oral-aural sections in first and second grades and were in LCE classes in third grade. The number of these pupils is shown below. Hereafter, these pupils will be referred to as members of the NOA-LCE group.

	Third Grade
NOA-LCE	59

The children who served as "controls" in the data analysis were pupils in the same schools as the experimental classes, who had never been in experimental classes, and for whom data were available. The table below reveals the number of children in the control group.

	Fourth Grade	Third Grade
Control	21	68

HYPOTHESES

The hypotheses for the fourth grade stated in null form follow:

1. After four continuous years the LCE, LCS and Control groups will not differ significantly on the following scores from the Iowa Tests of Basic Skills:

1.1 Vocabulary.

1.2 Reading Comprehension.

1.3 Language Skills--Spelling.

1.4 Language Skills--Capitalization.

1.5 Language Skills--Punctuation

1.6 Language Skills--Usage.

1.7 Language Skills--Total.

1.8 Work-Study Skills--Map Reading.

1.9 Work-Study Skills--Reading Graphs and Tables.

1.10 Work-Study Skills--Knowledge and Use of Reference Materials.

1.11 Work-Study Skills--Total.

1.12 Arithmetic Skills--Arithmetic Concepts.

1.13 Arithmetic Skills--Arithmetic Problem Solving.

1.14 Arithmetic Skills--Total.

1.15 Composite.

The hypotheses for third grade stated in null form follow:

2. After three continuous years the LCE, LCS, and NOA-LCE and Control groups will not differ significantly on the following scores from the Iowa Tests of Basic Skills:

2.1 Vocabulary.

2.2 Reading Comprehension.

2.3 Language Skills--Spelling.

2.4 Language Skills--Capitalization.

2.5 Language Skills--Punctuation.

2.6 Language Skills--Usage.

2.7 Language Skills--Total.

2.8 Work-Study Skills--Map Reading.

2.9 Work-Study Skills--Reading Graphs and Tables.

2.10 Work-Study Skills--Knowledge and Use of Reference Materials.

2.11 Work-Study Skills--Total.

2.12 Arithmetic Skills--Arithmetic Concepts.

2.13 Arithmetic Skills--Arithmetic Problem Solving.

2.14 Arithmetic Skills--Total.

2.15 Composite.

DESIGN AND METHOD OF ANALYSIS

Analyses were applied to each grade separately. First, a series of descriptive statistics was produced. The mean IQ and mean score on each test of the Iowa Tests of Basic Skills were computed for each experimental and control group along with the appropriate standard deviations. The statistical unit was the child, not the class, since there were no intact classes left.

Second, analysis of variance was used to compare the mean IQ and Iowa Tests of Basic Skills scores of the various groups at each grade level. Third, analysis of covariance was used to compare the Iowa Tests of Basic Skills scores for the groups using the IQ scores as the covariable.

ORGANIZATION OF REST OF THE STUDY

Chapter II contains a history of the San Antonio Language Research Project including a description of the instructional methods, rationales for those methods, and an explanation of administrative controls and sources of funds. Also included is a summary and critique of research pertaining to the Language Research Project. Chapter III relates in detail the procedures involved in the conduct of this study along with the findings. The summary, limitations, conclusions, and recommendations may be found in Chapter IV.

CHAPTER II

REVIEW OF PROJECT LITERATURE

HISTORICAL PERSPECTIVE

The San Antonio Language Research Project was inaugurated as Cooperative Research Project No. 2648 in the fall of 1964 with the cooperation of the San Antonio Independent School District, under the direction of Thomas D. Horn, chairman of the Department of Curriculum and Instruction, College of Education, The University of Texas at Austin. The Project was one of twenty-seven studies sponsored by the U.S. Office of Education in an effort to determine the best methods of teaching beginning reading.

The primary purpose of the Project in its first year was to "compare the effectiveness of three methods of developing reading readiness in Spanish-speaking boys and girls in the first grade."[1] The three methods were various types of oral language instruction that will be described later.

---

[1]Thomas D. Horn, "A Study of the Effects of Intensive Oral-Aural English Language Instruction, Oral-Aural Spanish Language Instruction and Non-Oral-Aural Instruction on Reading Readiness in Grade One," The University of Texas at Austin, 1966, pp. 3-4. (Hereafter referred to as "Year One Findings.")

Although the primary purpose of the first year was limited to development of reading readiness, the actual scope of the Project was much larger. According to Horn, the basic objectives were:

(1) To break through current methods and materials which have resulted in pupil failure, retention and ultimate school drop-out.
(2) To achieve a restructuring of teacher attitudes toward the Spanish-speaking school beginner which would foster a wholesome learning environment.
(3) To delineate in meaningful terms the forces affecting academic achievements by Spanish-speaking pupils, forces for which the schools may or may not be able to provide some form of compensatory education.
(4) To provide a reasonably defensible research base for using socio-economic and/or psychological data to predict the level of pupil achievement in school. If this is found possible, schools would be in a position to research compensatory educational programs in terms of their ability to overcome those factors which evidence a negative effect on school achievement.
(5) To obtain data which will enable defensible decisions to be made concerning (a) the role of oral language in the education of Spanish-speaking pupils; (b) the nature of "bi-lingualism" with particular reference to both positive and negative potentials for curriculum planning; (c) ways in which an educational program might be planned for Spanish-speaking pupils which would simultaneously develop cognitive and linguistic skills, using the basic content areas, e.g., science, social studies, mathematics, health and the like, as vehicles for language skill development; and (d) ways of enabling Spanish-speaking pupils to develop positive self-concept.
(6) To identify or develop diagnostic and achievement tests which are valid and reliable for the Spanish-speaking population.
(7) To develop more effective and pertinent in-service teacher education programs for teachers of Spanish-speaking pupils.[1]

---

[1] Ibid., pp. 7-8.

These objectives were maintained throughout the life of the Project, though limited resources prevented the fulfillment of each objective. The experimental treatments always consisted of oral language programs, but the original goal of developing reading readiness changed to one of improving reading skills.

The Project in Year One (1964-65) was supported by funds mainly from the U.S. Office of Education through the Cooperative Research Program and minimally from the Research and Development Center for Teacher Education, College of Education, The University of Texas at Austin. The Project was extended into Year Two (1965-66) and supported by funds from The University of Texas Research and Development Center for Teacher Education and by Elementary and Secondary Education Act (ESEA) Title I funds from the San Antonio Independent School District. In Year Three (1966-67) ESEA Title I funds again provided support while the R and D Center made possible the statistical analysis of data from the previous year and gathering of data for the current year.

In Year Four (1967-68) funds were provided through Title I and the Southwest Educational Development Laboratory (SEDL) in Austin. The R and D Center provided minimal funds for research purposes using existent data only. The data analyses for 1965-66 were finished and the data for 1966-67 and 1967-68 were collected and organized.

In Year Five (1968-69) financial support for the development and implementation of the special curriculum was provided by SEDL and Title I and III funds of the San Antonio Independent School District. No funds were available to The University of Texas at Austin for research purposes. The Project was extended into Year Six (1969-70) under the auspices of SEDL.

For the first three years (fall 1964--spring 1967) the experimental treatments consisted of three different oral language programs: intensive instruction in English, intensive instruction in Spanish and an approach using the same content but no intensive language instruction. The content of all three was based on lessons selected and adapted from Science: A Process Approach which is a series of exercises "designed to improve the child's skills in using the processes of science."[1] Thus, the goal was to teach not only oral language but also science skills such as observing, classifying and measuring. This involved using as much as possible of the "discovery approach" to

---

[1]Commission on Science Education, American Association for the Advancement of Science, Science: A Process Approach, Washington, D.C., 1963; Arthur H. Livermore, "The Process Approach of the AAAS Commission on Science Education," Readings in Science Education for the Elementary School, ed. by Edward Victor and Marjorie S. Lerner, The MacMillan Company, New York, 1967, p. 413.

teaching science. Also the pupils were to handle concrete objects during the activities. According to Arnold:

> This content was selected over other possible content areas because data informally analyzed had suggested that disadvantaged Spanish-speaking children did not find the content more difficult than did the more advantaged children. Only rarely will a child of any background have had extensive exposure to science concepts and the language of science before entering school. Also, the language of science tends to remain at the descriptive and objective level and is relatively free of the affective domain where differences in value systems and social systems may affect learning. Therefore, differences related to ethnic and socio-economic groups were assumed to be at a minimum.[1]

Another advantage of the AAAS materials and methods in this situation is that with or without the language teaching techniques the exercises are carried out without pupil books. The pupils need no reading skills to participate in the lessons.

The lessons used by the English and Spanish treatments have been described as being structured oral language lessons. Frequently this has been interpreted by observers to mean that the language lessons were built on some outline of the syntactic structures of each language. Actually, according to Ott, who was one of four authors of the

---

[1]Richard D. Arnold, "1965-66 (Year Two) Findings, San Antonio Language Research Project, Thomas D. Horn, Director," The University of Texas, Austin, 1968, p. 7.

original curriculum for the first and second grades[1] and who was the supervisor of the writers for the other grades, there was a concentrated effort at first to systematically include English sentence structures.[2] For the curriculum in third grade and beyond the writers were instructed to construct sentences necessary for the teachers and pupils to speak as they were to go through the selected exercises in the AAAS materials. The English language structures were not arranged in any particular order such as moving from the simple to the complex. The Spanish materials for all grade levels were developed by translating the English lessons. This procedure cannot be defended other than upon grounds of expediency.

The term "structured" referred to the fact that the teachers were given rather explicit directions for conducting the lessons. This was done deliberately, according to Ott, because when less "structured" lessons were tried out in summer school classes preceding their first use in the fall of 1964, there was great variation in teacher performance.

---

[1]David P. Butts, Elizabeth H. Ott, Albar A. Peña, and Anne O. Stemmler.

[2]Staff Seminar, Southwest Educational Development Laboratory, Austin, Texas, September, 1970.

The sentence patterns to be learned in both the intensive English and Spanish sections were presented or modeled by the teacher and repeated by the pupils in dialogue situations. After introduction of new patterns to the whole class, the teacher then initiated dialogues between teacher and small groups, between teacher and individual children, between groups of children and between individual children.

The dialogue in the lesson plans involved mostly questions and answers, affirmative and negative. Many of the answers involved elliptical responses, e.g., "Yes, they do." Some substitution drills were written into the lesson plans. The third and fourth grade teachers were expected to develop drills involving such transformations as the passive, e.g., "The dog bit the boy. The boy was bitten by the dog," and coordinate conjunction reduction, e.g., "The girl ran. The girl played. The girl ran and played." To the extent that the teachers were not able to produce their own drills, the oral language teaching techniques consisted only of modeling and repetition.

The emphasis in the lessons was on syntax and the use of complete sentences. Problems of phonology were treated informally and incidentally because standard English pronunciation was not considered one of the prime goals of the Project and because many of the teachers were unable to model standard pronunciation.

The group receiving intensive English instruction was called the OAE (oral-aural English) treatment for the first three years (1964-67). After that the name was changed to LCE (language-cognition English).

The Spanish group was called the OAS (oral-aural Spanish) treatment for three years and then became known as LCS (language-cognition Spanish). The Spanish treatment was originally included for two reasons:

(1) to test the effect of instruction in the students' native language on the development of oral English and subsequent reading in English; and
(2) to develop standard Mexican Spanish in place of limited local dialects.[1]

As subsequent information in this study will show, the effect of instruction in Spanish was not as beneficial as hoped. The second reason given above is regarded as linguistically naive. Regional Spanish dialects are not incomplete or unsystematic. They may be "different" from some other dialect, but they are not "limited."

The group using the science content but no intensive language practice was called the NOA (non-oral-aural) treatment. This group was originally intended to be a control group but actually became thought of as a third experimental treatment.

---

[1]Anne O. Stemmler, "An Experimental Approach to the Teaching of Oral Language and Reading," Harvard Educational Review, 36:45 (Winter, 1966).

> The major function of the NOA group was to provide data concerning possible "halo" effects, e.g., consultant attention and merely trying something different, caused by the Group 1 and 2 [OAE and OAS] activities and to determine the effect, if any, of the science materials alone on language and cognitive development and reading readiness.[1]

The NOA treatment was dropped after the first three years and the children were assigned insofar as possible to LCE and LCS classes.

Although the lessons were designed to be used only for oral language activities, it is known that at some point some teachers began to write the sentences from the special materials on charts or blackboards for the children to read. This change from the usual basal reading instruction was not an official component of the experimental treatments, but if it occurred more in one treatment than in another, it becomes a factor not accounted for by the research design.

Throughout the years of the Project the special lessons were supposed to be taught one hour per day for about 140 days. For the initial first grade sample (1964-65) language instruction received by the OAE and OAS groups replaced the regular reading readiness instruction. The NOA pupils received the special science instruction in addition to the regular reading readiness program of the

---

[1]Horn, "Year One Findings," p. 11.

school district. Thus, in the Year One study, when the three treatments were compared for their effectiveness as methods of developing reading readiness, the NOA group really received a combination of two methods of instruction: regular reading readiness and the informal language activities associated with the science activities.

Teachers in the experimental treatments received supervision and consultant help from The University of Texas at Austin or from the Southwest Educational Development Laboratory staff members. In-service sessions were intended to develop skills needed for oral language and science instruction and to increase understanding and acceptance of disadvantaged Mexican-American children. Teachers in the NOA treatment did not, of course, receive training in oral-aural teaching techniques.

The children who entered first grade in Year Two (1965-66) received basically the same type of instruction in the experimental treatments through Year Five (1968-69) as did the children who entered first grade in Year One (1964-65). The first, second and fourth grade materials were revised slightly and the third grade materials were completely rewritten before the second group of children were exposed to them.

Starting with Year Four (1967-68) a different set of treatments was begun which applied only to children who entered first grade that year and thereafter. The focus

of the research by The University of Texas at Austin faculty and students remained on children in the grades receiving the original treatments, however. The content of the new programs was expanded to include social studies as well as science. The lessons were taught in an ESL mode (English as a Second Language) or bilingually. There was neither a Spanish-only nor a non-oral-aural treatment. Some classes began the use of special reading materials which were designed to include only words and sentence patterns the children had learned previously in their language classes.

The original or slightly modified oral language materials based on science continued in use in ESL or bilingual classes in San Antonio and in other pilot test sites of the Southwest Educational Development Laboratory through the 1970-71 school year. The English version was used in Guam under Ott's supervision in the 1971-72 school year. Elsewhere a newly written set of materials was introduced by the Laboratory. The language exercises were very similar but the effort to develop science skills was discontinued and much of the science vocabulary was dropped.

YEAR ONE (1964-65) FINDINGS

As Director of Cooperative Research Project No. 2648, Horn reported the results of the research done during Year

One (1964-65).[1] His study will be referred to as "Year One Findings." His primary purpose was "to compare the effectiveness of three methods of developing reading readiness."[2] The three methods were the OAE, OAS and NOA treatments. There were nine classrooms receiving oral-aural instruction in English (OAE), ten classrooms receiving oral-aural instruction in Spanish (OAS), and nine receiving non-oral-aural instruction with the science exercises (NOA).

Horn also described a control group made up of children from twenty-one classrooms selected to represent all ethnic and socio-economic status groups in the San Antonio Independent School District. The control classes did not receive the special science curriculum in any form. Both the NOA treatment group and the control classes received the standard reading readiness program used in the San Antonio schools. Horn did not include data for the control group in his analysis.

---

[1]Thomas D. Horn, "A Study of the Effects of Intensive Oral-Aural English Language Instruction, Oral-Aural Spanish Language Instruction and Non-Oral-Aural Instruction on Reading Readiness in Grade One," The University of Texas at Austin, 1966. See also: Thomas D. Horn, "Three Methods of Developing Reading Readiness in Spanish-speaking Children in First Grade," The Reading Teacher, 20:38-42 (October, 1966).

[2]"Year One Findings," p. 3.

The OAE, OAS and NOA groups were reported by Horn to have received instruction from units based on a Texas Education Agency manual designed to serve as a guide to teachers in planning and teaching English to Spanish-speaking young children.[1] Apparently both the OAE and OAS teachers used lessons prepared by the Project staff. These lessons were in English and involved the use of audio-lingual techniques. Horn said that the NOA classes also used instructional units based on the TEA manual. How the NOA teachers used these materials and maintained the non-oral-aural nature of their treatment is not clear.

Horn did not report the amount of time spent by each treatment group on the adapted TEA materials. The more each group used the materials in English and with audio-lingual teaching techniques, the smaller the distinctions between the three experimental groups would have been. At this point, it can only be assumed that the science materials were the basis of the major part of the instruction, and the three treatments were indeed dissimilar.

Over 900 children were in the experimental classes. Complete initial data were obtained for 735 pupils. By the end of the year, transfers, drop-outs and absences decreased

---

[1]Texas Education Agency, Pre-School Instructional Program for Non-English Speaking Children, Bulletin 642, Texas Education Agency, Austin, 1964.

the number for whom complete data were collected to 584 pupils, including 316 boys and 268 girls. They were distributed as follows:

OAE = 186

OAS = 204

NOA = 194

Horn chose the class as the sample unit, rather than the pupil, in order to control in part the teacher variable. With the sample sizes as small as reported above, it is clear that although the class was the statistical unit, data were available for many fewer than the total number of children in at least some of the classes. (The classes were reported to have had about twenty-seven to thirty pupils.[1])

Because the Project was one of twenty-seven first grade reading studies sponsored by the U.S. Office of Education, the staff was expected to use several different instruments to assess pupil abilities. Many of the tests turned out to be completely inappropriate. Horn presented data demonstrating the inappropriateness of the following tests: Word Meaning and Listening subtests of the Metropolitan Readiness Tests (at the beginning of first grade), Phonemes and Letter Names sections of the Murphy-Durrell

---

[1]Conversation with Sue McClellan, a Project teacher, March 20, 1972.

Diagnostic Reading Readiness Test, Pattern Copying and Identical Forms.[1] In every case there were either large differences between local means and national means and/or a high number of zero scores. No appropriate test of oral language ability was available at that time, which left unmeasured the major thrust of the OAE and OAS treatments.

The analysis Horn provided used two written instruments: the Inter-American Test of General Ability administered in Spanish as a pretest in September and the Metropolitan Readiness Tests, Form A, administered in English as a posttest in the spring.[2] Although Horn did not say, the Metropolitan Readiness scores used must have been total scores. The Metropolitan Readiness Tests consist of these subtests: Word Meaning, Listening, Matching, Alphabet, Numbers and Copying. Similarly, the Inter-American Test of General Ability scores used in the analyses must have been

---

[1]Gertrude H. Hildreth, Nellie L. Griffiths, and Mary E. McGauvran Metropolitan Readiness Tests, Form A, Harcourt, Brace and World, Inc., New York, 1964; Helen A. Murphy and Donald D. Durrell, Murphy-Durrell Diagnostic Reading Readiness Test, revised edition. Harcourt, Brace and World, Inc., New York, 1964; Pattern Copying, research edition released by Thelma G. Thurstone; L.L. Thurstone and T.E. Jeffrey, Identical Forms, research edition, The Psychometric Laboratory, The University of North Carolina, Chapel Hill, 1956.

[2]Herschel T. Manuel, Prueba de Habilidad General, Serie Interamericana, Primario, Nivel 1, Forma DEs, research edition, Guidance Testing Associates, Austin, Texas, 1962.

total scores. This test consists of these subtests: Oral Vocabulary, Numbers, Association and Classification.

The first main question that Horn considered was: Will there be significant differences among the mean scores of the three groups? The means of the three groups on the posttest were:

Metropolitan Readiness, Form A[1]

Group	Means	N
OAE	48.19	9
OAS	49.86	10
NOA	54.86	9

These means can be compared to the national means for children entering first grade: 53.21.[2]

T-tests comparing each pair of means revealed no significant differences. This is interesting in light of the fact that significant differences were found in the pretest measure. The means for that measure were:

Inter-American Test of General Ability[3]

Group	Means	N
OAE	13.82	9
OAS	18.71	10
NOA	20.30	9

---

[1]Horn, "Year One Findings," p. 35.

[2]Ibid., p. 47.

[3]Ibid., p. 36.

According to t-tests, the differences between OAE and OAS and between OAE and NOA were both significant at the .02 level.

It was later discovered that some of the principals had assigned younger, less mature pupils and pupils who had recently arrived in the United States to the OAE group in hopes that they would receive special help. This violated the random assignment procedure called for by the experimental design and probably accounts for the lower OAE mean on the pretest.

Horn's third main question asked if there would be any differences among the mean posttest scores (Reading Readiness) for the three groups if the pretest scores (General Ability) were held constant. Because there was interaction among the posttest scores when controlling for pretest, the null-hypothesis associated with the third question was not tested.

The second main question compared the boys versus the girls instead of the treatment groups. For both pretests and posttests the differences between the means of the boys and the girls were non-significant (pretest: boys mean--17.77, girls mean--18.20; posttest: boys--49.91, girls--51.53). Interaction between the posttest scores when controlling for pretest prevented the testing of the null hypothesis that there would be no difference between

the readiness scores of the boys and girls when pretest scores are held constant.

Although not part of the original statistical design, Horn made several additional comparisons. He computed the posttest means for the girls and the boys in each class. Then he tested for significant differences between the highest and lowest means for: (1) all the girls; (2) the girls in each treatment group; (3) all the boys; and (4) the boys in each treatment group. Except for the girls in the NOA treatment group, all the differences were significant at the .01 level. This shows that there was a great range in initial ability and final readiness levels from class to class.

Similarly, when posttest means were figured for each class (boys and girls together) and comparisons were made of high and low means of all the groups and for each group separately, the differences were significant at the .01 level. Horn said that it was not clear whether this finding was due to teacher variables or to the existence of atypical classes.

Two major conclusions can be drawn from Horn's study. First, the experimental treatments did not produce differential effects in reading readiness. Therefore, this generalization adequately describes the outcome: NOA = LCE = LCS. Second, many standardized tests were

clearly inappropriate for the population in the San Antonio Language Research Project.

YEAR TWO (1965-66) FINDINGS

Arnold provided the results of the research done during Year Two (1965-66).[1] His report will be referred to as "Year Two Findings." The general problem he studied was whether or not differential growth occurred in reading achievement when children receive oral language training.

Arnold's study involved two samples: Sample I which consisted of second graders who had been studied as first graders by Horn and Sample II which consisted of first graders enrolled during 1965-66. The study of Sample II was intended to be a replication of Sample I. The special instruction received by Year Two's first graders was approximately the same as Year One's first graders. The major differences were that some of the language lessons had undergone some revision and many of the teachers were in their second year of experience with the new techniques. Most of the children in the second grade experimental classes had been in the same type of class the year before.

Arnold included in his data analyses the scores of a control group. These children were in classes "selected

---

[1]Richard D. Arnold, "1965-66 (Year Two) Findings, San Antonio Language Research Project, Thomas D. Horn, Director," The University of Texas at Austin, 1968.

from various schools in the district and represented a cross-section of socio-economic levels and ethnic groups, thus deviating from usual control methodology."[1]

Data from four types of tests were analyzed. These types and the specific tests are listed below:

1. Language Tests
   a. Brengelman-Manning Linguistic Capacity Index (first grade only.)[2]
2. Group Intelligence Tests
   a. IPAT Culture Fair Intelligence Test (second grade only).[3]
   b. Thurstone Pattern Copying Test (first grade only).[4]
   c. Goodenough-Harris Draw-A-Man Test (first grade only).[5]

---

[1] Ibid., p. 5.

[2] Frederick H. Brengelman and John C. Manning Linguistic Capacity Index, Fresno State College, Fresno, 1964.

[3] R.B. Cattell, Institute for Personality and Ability Testing (IPAT) Culture Fair Intelligence Test, Scale 1. Institute for Personality and Ability Testing, Champaign, Illinois, 1964.

[4] Pattern Copying, research edition released by Thelma G. Thurstone.

[5] F.L. Goodenough and Dale B. Harris, Goodenough-Harris Draw-A-Man Test, Harcourt, Brace and World, Inc., New York, 1963.

d. Test of General Ability, Inter-American Series (first grade only).[1]

e. Prueba de Habilidad General (first grade only).[2]

3. Reading Tests in English

a. Test of Reading, Inter-American Series.[3]

b. Metropolitan Achievement Tests (second grade only).[4]

c. Metropolitan Readiness Tests (first grade only).[5]

4. Reading Tests in Spanish

a. Prueba de Lectura, Serie Interamericana.[6]

---

[1]Herschel T. Manuel, Test of General Ability, Inter-American Series, Primary, Level 1, Form DE, Guidance Testing Associates, Austin, Texas, 1962.

[2]Herschel T. Manuel, Prueba de Habilidad General, Serie Interamericana, Primario, Nivel 1, Forma CEs, Guidance Testing Associates, Austin, Texas, 1962.

[3]Herschel T. Manuel, Test of Reading, Inter-American Series, Primary, Levels 1 and 2, Form DE, Guidance Testing Associates, Austin, Texas, 1962.

[4]Walter N. Durost (ed.), Harold N. Bixler, Gertrude H. Hildreth, and Kenneth W. Lund, Metropolitan Achievement Tests, Primary, Levels 1 and 2, Form A, Harcourt, Brace and World, New York, 1958.

[5]Gerturde H. Hildreth, Nellie L. Griffiths, and Mary E. McGauvran, Metropolitan Readiness Tests, Form A, Harcourt, Brace and World, New York, 1964.

[6]Herschel T. Manuel, Prueba de Lectura, Serie Interamericana, Primario, Niveles 1 and 2, Forma CEs, Guidance Testing Associates, Austin, Texas, 1962.

The control group was given only the Test of Reading, Inter-American Series, the IPAT Culture Fair Intelligence Test, the Metropolitan Achievement Tests, the Test of General Ability, Inter-American Series, and the Metropolitan Readiness Tests.

Given the large number of tests, it should not be surprising that obtaining complete data for all the children in the experimental and control classes was impossible.

The number of children for whom complete data were secured and the number of classes they represented is given below:[1]

Treatment Group	Grade Two		Grade One	
	Classes	Children	Classes	Children
OAE	12	178	12	187
NOA	11	177	11	160
OAS	10	178	7	105
Control	12	255	12	178
(Total with complete data:		788		630)
(Total enrollment:		1125		1050)

The purpose of the analysis was to compare the reading test scores of the three experimental groups and the control group.

---

[1]Arnold, "Year Two Findings," p. 8.

> Simple comparisons of means, however, were deemed inappropriate because of substantial initial differences in both reading-related skills and intelligence.[1]

The control groups' pretest scores were in all cases much higher than the experimental groups' scores. The NOA scores were generally second highest.

Arnold used analysis of covariance in order to equate the groups in terms of pretest scores. Because there was interaction between treatment and covariate (the regression lines were not parallel) in some cases, analysis of covariance was not always possible. In the situations where analysis of covariance was not possible and where the relationship between the dependent and independent variables was strong enough to warrant it, separate comparisons were made for high and low scores. Low scores were defined as those below the fifteenth percentile and high scores those above the eighty-fifth percentile.

The basic null hypothesis was that "the post-test scores most typically associated with a given pre-test score are equal for all treatments."[2] He outlined forty-one different null hypotheses. When differences between treatments were constant throughout the range of the

---

[1] Ibid., p. 13.

[2] Ibid., p. 22.

covariable, Arnold tested for differences between these groups at the second grade level:

(1) OAE vs. OAS.

(2) OAE and OAS pooled vs. NOA.

(3) NOA vs. Control.

(4) OAE, OAS and NOA pooled vs. Control.

He made comparisons of these groups at the first grade level:

(1) NOA vs. OAS.

(2) NOA vs. OAE.

(3) OAE vs. OAS.

(4) OAS vs. Control.

(5) OAE vs. Control.

(6) NOA vs. Control.

(7) OAE and OAS pooled vs. NOA.

(8) OAE and OAS pooled vs. Control

In the main body of the report Arnold first presented the means and standard deviations for each group at each grade level on each test and subtest. He also provided tables which summarized these analyses:

(1) Analysis of covariance of Test of Reading, Inter-American Series, Metropolitan Achievement Tests, and Prueba de Lectura, Serie Interamericana, Grade Two.

(2) Double covariance analyses of Test of Reading, Inter-American Series and Metropolitan Achievement Tests,

Grade Two. (The pretest scores of these tests and the IPAT Intelligence Test scores were the covariables.)

(3) Analysis of covariance of Prueba de Lectura, Serie Interamericana, Brengelman-Manning Linguistic Capacity Index, Test of Reading, Inter-American Series, and Metropolitan Readiness Tests, Grade One.

(4) Analysis of covariance of high and low pretest scores of Brengelman-Manning Linguistic Capacity Index, Test of Reading, Inter-American Series, and Metropolitan Readiness Tests, Grade One.

A lengthy appendix includes for each criterion and each appropriate covariate:

(1) the multiple correlation coefficient (squared) for the full model.
(2) the squared coefficient for the model obtained under the null hypothesis.
(3) the degrees of freedom associated with the F-statistic obtained by comparing the two models.
(4) the F-ratio.
(5) the probability of chance occurrence of full model values under the conditions imposed by the null hypothesis.[1]

Of special interest are the findings related to the Spanish reading test, Prueba de Lectura, Serie Interamericana. In making comparisons at both grade levels, when there was no interaction between treatment and covariate, Arnold found that the differences between groups in all cases were not significant. In other words, special oral

---

[1] Ibid., p. 79.

language instruction in Spanish had no differential effect on achievement in reading in Spanish--as measured by the Prueba de Lectura. Some teachers in the special language classes gave their children instruction in reading from experience charts using the sentences from the language lessons. If it had been known whether or not the children in OAS classes received this kind of instruction--reading in Spanish from experience charts--the data results would have been even more significant.

Because there were so many hypotheses investigated and so many tests used, the reporting of the results was long and complicated. The condensation here is taken from Arnold's summary.[1]

The results of the analysis of the data from Grade Two were fairly consistent. The differences between the OAE and OAS groups were significant only when the Word Knowledge subtest of the Metropolitan Achievement Tests was the criterion. This one difference favored the OAE group. All differences between the NOA group and the OAE and OAS groups together were significant, favoring the NOA group. When comparisons were made between the control group and the OAE, OAS and NOA groups together, the scores of the control group generally were significantly greater. It

---

[1] Ibid., pp. 64-66.

must be remembered that the composition of the control group was different from the experimental treatment groups, i.e., the children represented a cross-section of ethnic groups and socio-economic levels. In summary, this generalization is accurate for the second grade: Control > NOA > OAE = OAS.

The results of the analysis of the data from Grade One were not consistent. Moreoever, there was a marked tendency toward interaction between treatment and covariate; thus analysis of covariance was frequently inappropriate. Therefore, Arnold separately analyzed low and high pretest scores.

When comparisons of low pretest scores were made and when the criterion was the Test of Reading, Inter-American Series, the differences were non-significant or favored the OAE treatment. When the criterion was the Metropolitan Readiness Tests, the control group scores were generally significantly higher.

When comparisons of high pretest scores were made and when the criterion was the Test of Reading, Inter-American Series, the OAE group was significantly superior to the NOA and OAS groups. Differences between the OAE and control groups were not significant. When the criterion was the Metropolitan Readiness Tests, there were no significant differences between each pair of the

experimental groups OAE, OAS and NOA. The differences between the control group and each experimental group were either not significant or favored the control group.

In summary, the findings for the first grade showed that in general when the criterion was a reading test, the OAE treatment was superior but when it was a readiness test, the control group was superior.

YEAR THREE (1966-67) FINDINGS

Knight presented the results of the research done during Year Three (1966-67).[1] Knight's report will be referred to as "Year Three Findings." As with Arnold's study of the Year Two (1965-66) findings, the major problem was to evaluate the effect of oral language training on reading achievement. In studying this problem, he asked three questions:

(1) Is one treatment generally more effective than the other treatments?

(2) Do children who score high, middle or low in the fall tend to perform similarly in the spring also?

(3) Does the difference between treatments vary according to pretest scores?

---

[1]Lester N. Knight, "A Comparison of the Effectiveness of Intensive Oral-Aural English Instruction, Intensive Oral-Aural Spanish Instruction, and Non-Oral-Aural Instruction on the Reading Achievement of Spanish-Speaking Second- and Third-Grade Pupils," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1969.

The groups were the same OAE, OAS and NOA groups studied by Horn and Arnold. Knight used no control group.

According to the description procedures in his summary chapter, Knight randomly selected thirty pupils in one treatment group for the data analysis.[1] Then he chose thirty pupils from each of the other two groups who had pretest scores identical with those in the initial group.

Knight used one group intelligence test, two English reading tests and one Spanish reading test. The group intelligence test was given in the fall to both grades: IPAT Culture Fair Intelligence Test (3rd grade, Scale 2; 2nd grade, Scale 1).[2] The reading tests were given to both grades in the fall and spring: Metropolitan Achievement Tests (Primary I, Form A; Primary II, Forms A and C); Test of Reading, Inter-American Series (Primary, Level 1, Form DE, and Primary, Level 2, Forms CF and DE); and Prueba de Lectura, Serie Interamericana (Primario, Nivel 1, Forma CEs and Primario, Nivel 2, Formas CEs and DEs).

Knight classified the fall scores of his samples at both grade levels in two dimensions: treatment and level of pretests (high, medium and low). He used spring scores as dependent variables and "conducted two-factor analyses of variance to obtain F ratios and tests of

---

[1] Ibid., p. 201.

[2] See the previous section for publication information for all the tests listed here.

significance for treatment main effect, pretest main effect, and interaction."[1]

Knight tested six null hypotheses for each grade level. The pretest and posttest scores involved in testing each hypothesis are listed below:

Hypothesis	Pretest Scores	Posttest Scores
1	IPAT Culture Fair Intelligence Test	Metropolitan Achievement Tests
2	IPAT Culture Fair Intelligence Test	Test of Reading
3	IPAT Culture Fair Intelligence Test	Prueba de Lectura
4	Metropolitan Achievement Tests	Metropolitan Achievement Tests
5	Test of Reading	Test of Reading
6	Prueba de Lectura	Prueba de Lectura

Each of the spring tests had three subtest scores and a total score, so for each hypothesis there were four differences to be tested. Thus there were twenty-four hypotheses all together.

In answer to his first question (treatment main effect), Knight found that fifteen of the twenty-four possible differences at the third grade level were significant. Nine favored the NOA group. Six favored the OAE group. For second grade, only seven of the treatment findings were significant. Six favored the OAE group and one the NOA group.[2]

---

[1]Knight, "Year Three Findings," p. 128.

[2]Ibid., pp. 205, 208 and 209.

In answer to the second question (pretest main effect), he found that the pretest effect was highly significant in almost all of the analyses. In other words, children who scored high, middle or low in the fall tended to perform similarly in the spring.[1]

In answer to the third question (interaction), he found only a few significant interaction effects and they followed no consistent pattern. This means that if a certain treatment resulted in the highest scores at the high pretest level, it tended to do the same for the other levels of the pretest.[2]

In summary, none of the significant differences regarding treatment favored the OAS group, even on the Spanish reading test, Prueba de Lectura. The significant treatment findings for third grade tended to favor the NOA group; the significant treatment findings for second grade tended to favor the OAE group. This latter might be explained by the fact that most of the second grade teachers had worked with the new materials and methods the year before.

The results of research performed during Year Four (1967-68) are reported in Chapter III of this study.

---

[1] Ibid., p. 202.

[2] Ibid., p. 214.

YEAR FIVE (1968-69) FINDINGS

Taylor reported the results of the research done during Year Five (1968-69).[1] Her dissertation will be referred to as "Year Five Findings." She developed an instrument to measure "quality and quantity of productive language" in order to determine if the pupils exposed to the LCE and LCS treatments for four or five years had benefitted significantly in language development.[2] She compared the language of the experimental groups of pupils with that of the "control" children, who had received no special language treatment, with regard to phonology, intonation and fluency.

Taylor made an effort to identify all pupils still in Project classrooms who had received the same treatment during their entire school career. Table 1 shows the number of pupils identified as being in continuous treatment compared to the number of children for whom complete data were collected in the first year. The original Project involved 28 classrooms in nine schools. By Year Five (1968-69) the students in continuous treatment were

---

[1]Thomasine H. Taylor, "The Effects of Oral-Aural Language Training on Gains in English Language for Fourth and Fifth Grade Disadvantaged Mexican-American Children," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1969.

[2]Ibid., p. 8.

Table 1

Number of Children Remaining in Original Treatment Groups in Year Five (1968-69) Compared to Number for Whom Complete Data Were Collected in First Grade, San Antonio Language Research Project[1]

Treatment	Grade 1 (1964-65)	Grade 5 (1968-69)	Grade 1 (1965-66)	Grade 4 (1968-69)
	Sample I		Sample II	
LCE[2]	186	34	187	125
LCS[3]	204	33	105	29

[1]Compiled from data presented by Taylor, op. cit., p. 63; Horn, op. cit., p. 18; and Arnold, op. cit., p. 18.

[2]LCE = Language-Cognition English

[3]LCS = Language-Cognition Spanish

in only six schools. Taylor did not say how many different teachers, i.e., classrooms, were involved.

The number of pupils who received continuous treatment dropped considerably except for the fourth grade LCE group. Taylor did sample randomly, however, from that group so that for her study the size was reduced to twenty-seven. The children in the control groups were selected from a school not in the Project but which had a population similar to that of the schools in the Project. Due to absences, transfers and inaudible recordings of some of the oral responses, the number of children in each treatment group and the number of controls were reduced further. The final number of children for whom data were collected and analyzed is shown below:

	Sample I (Fifth Grade)	Sample II (Fourth Grade)
LCE	32	27
LCS	26	20
Control	29	27

In order to test oral language proficiency, Taylor chose to take measures of phonology, intonation and fluency. She did this by using two filmstrips and recorded modeled sentences from two sets of Gloria and David Series materials used for teaching English or Spanish as a second language. Test 6 contained twenty pictures which the

children watched as forty sentences were modeled on a tape.[1] The children were asked to say the sentences "exactly the way" the lady on the tape said them. The children's sentences were recorded and evaluated later for both phonology and intonation. For testing fluency, five frames of another filmstrip were shown to the children.[2] Each frame was shown for thirty seconds and the children were asked to pretend that they were radio announcers telling people about the pictures. They were requested to keep speaking until a new picture appeared.

To measure phonology Taylor evaluated the last thirty-six of the forty sentences repeated on the tape accompanying the first filmstrip. (The first four responses were regarded as practice sentences.) She counted as errors only gross deviations which would be apparent to a lay listener like a classroom teacher. To get a total score, the number of deviations was subtracted from 616, the number of phonemes in the thirty-six sentences.

On intonation, four points were possible for each sentence. Therefore, a total of 144 points was possible. Taylor gave two accounts for how the maximum number of

---

[1]Gloria and David Beginning English, Series No. 20, Test 6, Language Arts, Inc., Austin, Texas, 1958.

[2]Gloria and David Spanish Intermediate, Series No. 60, "Saturday's Activities," Language Arts, Inc., Austin, Texas, 1958.

points could be earned. According to one, four points were given for a close approximation to the modeled sentence. In the other explanation, she said that if it was obvious that the child knew what he was saying, four points were awarded. In a personal communication Taylor said that these two criteria did not conflict.

On intonation, a child was given three points if he repeated the sentence with meaningful intonation but missed special stress on a particular word. If the child garbled the sentence or appeared to be trying to remember only isolated words, he received two points. If there was no response, one point was given. From the description of the criteria Taylor used to judge the responses, it is not clear what her definition of intonation was.

To score fluency, Taylor counted the words said in response to the pictures on the second filmstrip. She provided no definition of fluency and offered no rationale which explained why a word count was chosen as the measure of fluency.

Taylor also computed a total language score by adding the other three scores together. Adding together raw scores to get a total score may be a questionale procedure.

The following hypotheses were considered in the analysis of the data:

There will be no significant difference among the LCE, LCS and Control groups on Phonology, Intonation, Fluency and Total Language at the fifth grade and at the fourth grade levels (Hypotheses 1-8).

There will be no correlation between the three subscales and total score: (a) Phonology and Intonation; (b) Phonology and Fluency; (c) Intonation and Fluency; and (d) each subscale with Total Language (Hypothesis 9).

Analysis of variance was performed to test the first eight hypotheses. Table 2, an adaptation of one of Taylor's tables, shows the order of the means for each group for Phonology, Intonation, Fluency and Total Language at each grade level. Included also is an indication of whether or not the differences in the means reached significance.

For Sample I (Grade Five) analysis of variance showed that the Phonology scores were significantly different at the .01 level and the Fluency and Total Language scores were significantly different at the .05 level. The LCE treatment group received the highest mean score in Phonology and the LCS treatment group received the highest mean scores in Fluency and Total Language. In all three cases, the control group received the lowest mean scores.

Table 2

Summary of Order of Means for Phonology, Intonation, Fluency, and Total Language, Year Five (1968-69), San Antonio Language Research Project[1]

	Sample I (Grade 5)	Sample II (Grade 4)
Phonology	LCE[2] > LCS[3] > Control**	LCS > LCE > Control
Intonation	LCS > Control > LCE	LCE > Control > LCS
Fluency	LCS > LCE > Control*	LCS > Control > LCE
Total Language	LCS > LCE > Control*	LCS > Control > LCE

* $p < .05$
** $p < .01$

[1]Compiled from Taylor's Table 16, "Year Five Findings," p. 110.

[2]LCE = Language-Cognition English

[3]LCS = Language-Cognition Spanish

For Sample II (Grade Four) none of the findings were significant. The LCS treatment group scores were highest on all tests except Intonation.

Taylor computed an intercorrelation matrix to test the ninth hypothesis using the three subtest scores and total scores for children at both grade levels. Her Table 9 is reproduced as Table 3. It should be noted that although all the correlations were high enough to be significantly different from zero, they were not necessarily high enough to account for a very important proportion of the variance in one score which can be attributed to variation in the other.

The correlation between Total Language and Fluency is so high (.96) that, although the correlation is spuriously high because the Total Language score includes the Fluency score, Taylor is justified in concluding that the Fluency test gives essentially the same information as the whole series of tests.[1] She may not be justified, however, in suggesting also that the Fluency test be used as a "quick measure of general language proficiency" because she has not provided an argument showing that her test does in fact measure "general language proficiency."[2]

---

[1]Quinn McNemar, Psychological Statistics, 3rd. ed., John Wiley and Sons, Inc., New York, 1962, p. 164.

[2]Taylor, "Year Five Findings," p. 119.

Table 3

Correlation Coefficients Computed on All Students in the Sample (N=161) for Phonology, Intonation, Fluency, and Total Language, Year Five (1968-69), San Antonio Language Research Project[1]

	Phonology	Intonation	Fluency
Intonation	.2213**	----	----
Fluency	.4198**	.1600*	----
Total Language	.6030**	.2870**	.9640**

* $p < .05$
** $p < .01$

[1]Taken from Taylor's Table 9, "Year Five Findings," p. 98.

Although Taylor referred to her test as one of general language proficiency, the test does not measure control of syntax nor the extent of each child's vocabulary. In addition, the group means for both grades on Intonation, ranging between 141.05 and 142.31, were so close together and so close to the total possible score of 144 that one may conclude that the measure of intonation was not very sensitive. An alternative explanation would be that intonation was quite good for all groups. Taylor did speculate in a discussion of the limitations of her study that the scoring system was inappropriate and that a better one would involve only a right or wrong rating.

Observation of the data prompted Taylor to add one comparison to her analysis. She compared the Grade Five scores on all subtests and the total with the Grade Four scores for each treatment group and the control group. Disregarding Intonation scores which have the difficulty previously mentioned, the Grade Five scores were higher than the Grade Four scores in all cases except on the Fluency subtest and the Total Language score for the control groups. Taylor used a t-test to see if the Fluency scores were significantly different and found that they were not. She did not determine if the Total Language scores were also significantly different.

In discussing these findings, Taylor commented in a misleading way. Since no data were supplied, nor available, concerning the starting points of the children in the fifth or fourth grades as they began their school careers or at the beginning of the Year Five school year (1968-69), it might have been more accurate for Taylor to discuss the differences between the fourth and fifth grade control scores rather than referring to a regression.

RELATED STUDIES

Studies of Oral Language

Ott

Ott's dissertation is one of several reporting the development and use of instruments for assessing oral language,[1] and will be referred to as "A Study of Oral English." She stated that the intent of her study was "to determine levels of fluency and proficiency in the use of oral English, expressive operant, of Spanish-speaking school beginners, comparing two methods of instruction."[2] The two methods were the OAE and NOA treatments for Grade One students (Sample II) in Year Two (1965-66). She went on

---

[1]Elizabeth H. Ott, "A Study of Levels of Fluency and Proficiency in Oral English of Spanish-Speaking School Beginners," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1967.

[2]Ibid., p. 8.

to say, however, that besides comparing the ability to speak (expressive operant) of the OAE and NOA treatment groups, she was also going to compare the ability of the groups "to communicate in oral English in the receptive operant"[1] which she defined as the ability to "understand verbal instructions and to comprehend language used in basic everyday communication in the school environment when spoken at a conversational rate."[2]

Ott used the Brengelman-Manning Linguistic Capacity Index to assess receptive language abilities (understanding).[3] To measure productive language abilities (speaking) she collaborated with Jameson on the development of a new instrument.[4] This test was referred to by several different names in both the Ott and Jameson dissertations. Here it will be called the Ott-Jameson Test.

The Linguistic Capacity Index is a group test designed for use with primary grade pupils whose native language is Spanish. The examiner asks the children to mark or circle the appropriate picture in a set of line

---

[1] Ibid., p. 9.

[2] Ibid., p. 12.

[3] Frederick H. Brengelman and John C. Manning, Linguistic Capacity Index, Fresno State College, Fresno, 1964, Manual (ditto), Pupil Booklet (offset).

[4] Gloria R. Jameson, "The Development of a Phonemic Analysis for an Oral English Proficiency Test for Spanish-Speaking School Beginners," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1967.

drawings. By marking the right picture the children indicate their understanding of certain vocabulary items, sound contrasts and grammatical structures. The three sections of the test, Vocabulary Recognition, Contrastive Phonology, and Contrastive Grammar, each contain twenty items.

Ott began development of the Ott-Jameson Test in July, 1965. It was administered in a pretrial run to twenty subjects during the first two weeks of September. It was then administered to seventy-five first grade subjects.

At the time of the September administration, the test consisted of three parts. In the first part the examiner showed drawings to each child and spoke six to nine sentences for each of twenty-five phonemic problems. The child was told to repeat each sentence. His response was tape recorded in order to be evaluated later. In the second part the child was asked to select three items from a group of five: doll dress, whistle, toy automobile, plastic wristwatch and a small cream pitcher. The examiner then asked five questions about each item. In the third part, the examiner showed the subject three photographs (children eating and drinking, children playing on a playground, and children playing in groups) and asked him questions about each photograph.

After the test was administered in September, Ott decided to eliminate the second part since it elicited a limited number of measurable responses. Jameson was asked to revise the first part and to devise a scoring system for it. Jameson's work on the test is further described in the next section of this paper. According to Ott, the first part was changed only by eliminating one phonemic problem and adding one other.[1] This seems to be contradicted by Jameson, who reported making three separate revisions between the September and May administrations of the test.[2] Since Ott did not provide a copy of the September version, no direct examination of the two versions was possible. Ott did not include the three items related to the phonemic problem added to the May version in the pre-post comparisons. The sentences and questions for the two remaining parts to be used in May were recorded on a master tape; the photographs for the second part were replaced by color filmstrips.

Ott included results of fifty-four sentences in the first part, the phonemic analysis. Each sentence contained one word which was judged as: (1) pronounced correctly; (2) pronounced with a minimum phonemic change, i.e., with

---

[1]Ott, "A Study of Oral English," p. 65.

[2]Jameson, op. cit., pp. 19-23.

the predicted error; or (3) not pronounced at all, pronounced with an error not predicted or pronounced with a response not identifiable. Pronunciation of the other words in each sentence was not relevant for the scoring scheme.

The second part of the Ott-Jameson Test included twenty-nine questions about several pictures. The questions were classified as being on the literal level, the non-literal/inferential level or the imaginative level. This classification was not taken into account in the scoring system, however. The scoring was done by counting: (1) the total number of responses; (2) the number of responses using Spanish intonation; (3) total word count with Spanish intonation; (4) the number of responses using English intonation; and (5) total word count with English intonation.

Apparently, if a child answered a question by one word, a phrase, a sentence, or several sentences, he was given credit for one "response." Since intonation refers to stress and pitch contours, it is not clear how the term intonation was used with reference to individual words in the word count measures.

In various places, Ott said that the second part of the test was intended to measure:

(1) appropriate word choice (or vocabulary);

(2) control of syntax (or structural features);

(3) intonation;

(4) juncture;

(5) stress;

(6) extent of prompt and fluent response as measured with a word count; and

(7) fluency as evidenced by extent of prompt and continuous expression in English.[1]

It is not obvious how the method of scoring actually measured all these aspects of language. In no place did Ott explain the relationship between her scoring system and what the test was intended to measure. One wishes that Ott had more clearly spelled out the procedures used in scoring the tapes for the responses to the Ott-Jameson Test. For instance, it is important to know if the scorer was aware of the treatment group from which each subject came before ruling out the possibility of systematic bias in the data analyses.

In addition to administering the two language tests, Ott also collected the IQ scores from the fall administration of the Goodenough-Harris Draw-A-Man Test.[2]

---

[1]Ott, "A Study of Oral English," pp. 10, 11, 22, 118, 119, and 141.

[2]F.L. Goodenough and Dale B. Harris, Goodenough-Harris Draw-A-Man Test, Harcourt, Brace and World, Inc., New York, 1963.

Ott chose a small sample from four of the eight Project schools "on the basis of convenience."[1] Seventy-five children were tested in September and fifty-eight in May. Every child met the following criteria:

(1) Score no lower than 70 on the Goodenough-Harris Draw-A-Man Test.

(2) Age no greater than 7.0 as of September 1, 1965, and no previous exposure to a first grade instructional program.

(3) Little or no functional use of oral English as indicated by level of fluency and proficiency as determined by the Ott-Jameson Test.

(4) Present in class and physically well enough to participate in pre- and post-testing.[2]

The rationale for the third criterion listed above is not apparent. Moreover, one does not know how a subject will perform until after the test has been administered, so the third criterion must have been applied ex post facto. Application of the third criterion appears to mean that children with initial strength in English were excluded from the sample. Thus, any conclusions

---

[1]Ott, "A Study of Oral English," p. 115.

[2]Ibid., p. 19.

reached about the effectiveness of the OAE and NOA treatments apply only to children who initially had "little or no functional use of oral English."

The fifty-eight children in the final sample were evenly distributed between the NOA and OAE treatment groups. In light of Ott's own conclusions and of later findings, it was unfortunate that she did not include also children from the OAS treatment group and from classes receiving instruction in the regular San Antonio curriculum.

In her data analysis, Ott determined correlations for the variables listed below. These were also the variables to which she applied analysis of variance.

Variable	Name
1	Correct Response (Ott-Jameson, Part I)
2	Minimum Change (Ott-Jameson, Part I)
3	Missed (Ott-Jameson, Part I)
4	Total Responses (Ott-Jameson, Part II)
5	Responses, Spanish Intonation (Ott-Jameson, Part II)
6	Word Count, Spanish Intonation (Ott-Jameson, Part II)
7	Responses, English Intonation (Ott-Jameson, Part II)
8	Word Count, English Intonation (Ott-Jameson, Part II)

9 Vocabulary Recognition (Linquistic Capacity Index)

10 Contrastive Phonology (Linquistic Capacity Index)

11 Contrastive Grammar (Linquistic Capacity Index)

12 Total Score (Linquistic Capacity Index)

13 Goodenough-Harris IQ Score (Draw-A-Man Test)

Ott first provided a table showing the Pearson Product Moment correlations computed for the September and May scores of the language tests. These test-retest correlations were presented as evidence of reliability of the tests but the conclusions must be questioned because of the time lapse between September and May, during which the ability being measured may have changed. In fact, the goal of the experimental program was to enhance the very abilities being tested. Therefore the reliability coefficients were not the correlations "between two comparable measures of the same thing."[1] Moreover, reliability of a test should be established on groups different from the one for which the test is being used as a measure of growth.

Ott also included three tables which gave Pearson Product Intercorrelation Matrices for the total sample, the

---

[1]McNemar, op. cit., p. 146.

OAE group and the NOA group. The most interesting finding was that for the total sample all correlations between the Ott-Jameson Test variables and IQ at both September and May administrations failed to be significant at the .05 level, i.e., $r > .23$. The highest correlation with IQ of a Linquistic Capacity Index variable was only .28 for Contrastive Phonology in September.

A third set of tables showed means and standard deviations for September and May and the mean gains for each variable for the OAE group and the NOA group. The tables contained data for an extra variable: Total Word Count. Ott gave no explanation for this addition nor for how the Total Word Count could be less than the sum of the Word Count, English Intonation and the Word Count, Spanish Intonation. A second item not explained was a standard deviation of 37.99 for Word Count, English Intonation in September for the OAE group while the mean was only 21.48.

For the OAE group the trends easily seen on the table of means and standard deviations included a large increase in Correct Responses on the phonology part of the Ott-Jameson Test, and in Responses, English Intonation and Word Count, English Intonation on the fluency part of the Ott-Jameson Test. Thus, there were also large decreases in the number of Minimum Change responses, Responses, Spanish Intonation, and Word Count, Spanish Intonation.

For the NOA group there were large increases in the number of Correct Responses on the phonology section of the Ott-Jameson Test and in Word Count, English Intonation. There were no large decreases. Even the Word Count, Spanish Intonation was greater in May.

When one compares the word counts to the number of responses of a particular intonation for both groups, it becomes obvious that the children must have spoken in sentences of from an average of two and a half to five words. The shortness of the sentences was neither commented on nor explained by Ott.

Ott then used repeated measures analysis of variance on each variable in the language tests except for the Total Score on the Linguistic Capacity Index. She provided tables and figures for each variable giving F's and p's for:

(1) groups;

(2) testing periods; and

(3) groups across testing periods.

In other words, she compared:

(1) the combined pretest and posttest scores of the two groups (OAE and NOA);

(2) the combined OAE and NOA scores at the two administration times (May and September); and

(3) the gains in the scores of the two groups between September and May.

Table 4 summarizes these comparisons.

All the scores between testing periods for the groups together (comparison #2 listed above) differed significantly at the .001 level except for Variable 6, Word Count, Spanish Intonation, which differed significantly at the .03 level. All changes from September to May for both were in directions indicating growth in English oral language as measured by the Ott-Jameson Test and the Linguistic Capacity Index except for one. The NOA group increased slightly on Word Count, Spanish Intonation.

The scores of the two groups (OAE and NOA), collapsing over time (comparison #1 listed above), differed significantly only on Variable 2, Minimum Change ($p = .04$), Variable 5, Responses, Spanish Intonation ($p = .02$) and Variable 8, Word Count, English Intonation ($p = .01$).

On Variable 2, Minimum Change, the OAE group was nine responses higher than the NOA group in September. Both groups showed a decrease in Minimum Change responses between September and May, but the OAE group had a much greater decrease (a decrease indicates improvement in standard English pronunciation). The difference in gain scores for the two groups was significant at the .001 level (comparison #3 listed above).

The OAE and NOA groups both scored thirteen on Variable 5, Responses, Spanish Intonation, in September but in May the OAE group dropped to five while the NOA

Table 4

F-Ratios and Probability Levels for Three Comparisons of Ott-Jameson Test and Linguistic Capacity Index Variables for First Graders in OAE (Oral-Aural English) and NOA (Non-Oral-Aural) Treatment Groups, Year Two (1965-66), San Antonio[1]

Variable	Groups F	Groups p	Testing Periods F	Testing Periods p	Groups Across Testing Periods F	Groups Across Testing Periods p
Ott-Jameson Test						
1 Correct Response	< 1.00	n.s.	369.32	<.001	68.16	<.001
2 Minimum Change	4.38	.04	290.71	.001	76.58	.001
3 Missed	< 1.0	n.s.	14.28	.001	< 1.0	n.s.
4 Total Responses	< 1.0	n.s.	81.22	.001	9.03	.005
5 Responses, Spanish Intonation	5.61	.02	40.09	.001	21.02	.001
6 Word Count, Spanish Intonation	1.79	>.18	5.41	.03	15.94	.001
7 Responses, English Intonation	2.73	>.09	328.95	.001	85.91	.001
8 Word Count, English Intonation	7.82	.01	54.14	.001	26.56	.001
Linguistic Capacity Index						
9 Vocabulary Recognition	1.19	>.20	96.18	.001	< 1.0	n.s.
10 Contrastive Phonology	3.05	<.08	70.19	.001	< 1.0	n.s.
11 Contrastive Grammar	< 1.0	n.s.	116.71	.001	< 1.0	n.s.

[1]Compiled from Ott's Figures 1-11, "A Study of Oral English," pp. 81, 83, 84, 86, 88, 89, 91, 93, 95, 96 and 97. No explanation was given as to why inequality signs for F's and P's were used in some tables and not in others.

group dropped only about two responses. Again the difference in gain scores between the OAE and NOA groups was significant at the .001 level (comparison #3 listed above).

On Variable 8, Word Count, English Intonation, the two groups scored almost the same in September (twenty-one), but the OAE group increased to 112 in May; the NOA group increased only to thirty-seven. The difference in gain scores was significant at the .001 level (comparison #3 listed above).

These data for the three variables (2, 5 and 8) all support the contention that the experimental program results in greater growth in oral English than unstructured language activities of the non-oral-aural approach.

The gain scores (comparison #3 listed above) differed significantly also for Variable 1, Correct Responses ($p = .001$), Variable 4, Total Responses ($p = .005$), Variable 6, Word Count, Spanish Intonation ($p = .001$) and Variable 7, Responses, English Intonation ($p = .001$). In all four cases, the gains were in directions favoring the OAE group.

Notice finally that on the Ott-Jameson Test, the differences between the gains of the two groups (comparison #3 listed above) were significant for all variables except Variable 3 (Missed--on the phonology section), while none of the Linguistic Capacity Index gain scores was significantly different.

After presenting the results of applying analysis of variance to the variables, Ott used the findings for some of the variables to test six hypotheses.

The first three hypotheses were concerned with expressive language.

Hypothesis 1: There will be no difference between treatment groups in proficiency in language as indicated by the ability to reproduce English phonemes accurately from a model. On the basis of the results of analysis of Variables 1 and 2, this null hypothesis was rejected.

Hypothesis 2: There will be no difference between treatment groups in proficiency in the use of English, including appropriate word choice, control of syntax, intonation and juncture. Using the results of the comparisons on Variable 7, she rejected the null hypothesis. Given that Variable 7 was derived by counting the number of responses using English Intonation to the twenty-nine questions in the second part of the Ott-Jameson Test, it is a little difficult to believe that Hypothesis 2 was completely tested.

Hypothesis 3: There will be no differences between treatment groups in the extent of prompt and fluent response in English. This was measured by Variable 8, Word Count, English Intonation. Again the null hypothesis was rejected.

The second three hypotheses were concerned with expressive language.

Hypothesis 4: There will be no significant difference between treatment groups in number of correct responses on the Vocabulary Recognition section of the Linguistic Capacity Index. The findings related to Variable 9 revealed that Hypothesis 4 could not be rejected.

Hypothesis 5: There will be no difference between the treatment groups in number of correct responses on the Contrastive Phonology section of the Linguistic Capacity Index. Comparisons on Variable 9 did not permit the hypothesis to be rejected.

Hypothesis 6: There will be no difference between the treatment groups in number of correct responses on the Contrastive Grammar section of the Linguistic Capacity Index. Comparisons on Variable 10 did not permit the hypothesis to be rejected.

In summary, both the OAE and NOA groups achieved significant gains in expressive language as measured by the Ott-Jameson Test and in receptive language as measured by the Linguistic Capacity Index. The gains for the OAE group were significantly greater than those for the NOA group in expressive language, while there were no differences between the groups in amount of gain in receptive language.

Ott felt that the learning experiences provided in the science program used by the NOA group were an "important factor in maintaining a consistent and equal gain with OAE

in language proficiency in the Receptive Operant."[1] Since Ott's comparisons did not include groups which did not receive the science curriculum, however, she did not have evidence showing that the "regular" curriculum and its "traditional" language activities would be any less effective in producing gains in receptive language abilities.

Jameson

Jameson developed an instrument for testing the ability of children to produce certain English phonemes.[2] Her dissertation will be referred to as "Phonemic Analysis." Her test was prepared with four objectives in mind:

(1) to obtain further information on the basic language problems of disadvantaged preliterate Spanish-speaking children in San Antonio public schools;

(2) to develop an effective test for assessing the oral English level of these children;

(3) to develop a test that could be used by a classroom teacher with no linguistic background, following brief training;

(4) to develop a test that could measure the progress of the child in his discrimination and reproduction of oral English, between the periods in which the test was administered.[3]

---

[1]Ott, "A Study of Oral English," p. 122.

[2]Gloria R. Jameson, "The Development of a Phonemic Analysis for an Oral English Proficiency Test for Spanish-Speaking School Beginners," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1967.

[3]Ibid., p. 5.

These objectives can be collapsed into one: to prepare a test simple enough for teachers with brief training to administer which could be used as a pretest for use in diagnosing needs of preliterate children and as a posttest for measuring progress.

In order to develop the instrument and to meet her objectives, Jameson found it necessary to:

(1) Review the literature on English tests for non-native speakers;

(2) Review the literature on phonological analyses of English and Spanish;

(3) Write an outline of a contrastive analysis of English and Spanish;

(4) Make three revisions of Part I of an oral language test first developed by Ott;

(5) Administer each revision of the test;

(6) Make a phonetic and a phonemic transcription of twelve of the tapes of the third revision (the phonemic description was used by Jameson in writing the summary statements for the data analyses);

(7) Compare her method of scoring the tests with scoring done by classroom teachers; and

(8) Analyze and summarize the findings from the administration of the third revision.

Ott's version of the test was administered to seventy first grade children (Sample II) in Project schools in September, 1965 (Year Two).[1] It "used eight items (groups of pictures) to elicit the target response words in the phonemic problem areas."[2] Ott began a revision of this test in October but Jameson took over the responsibility for the revisions in November. Ott continued to work on a test for other aspects of productive language.

Jameson's first revision consisted of eighty-seven short minimal pair sentences which checked twenty-one phonemic problems "that had been reported by teachers, or were predicted for the native Spanish-speaking child."[3] Colored pictures illustrating each sentence were shown to the child as he listened to a model saying the sentence. The child was to repeat the sentence. Scoring was done later by playing back the tape on which each child's voice had been recorded.

The first revision was administered in January, 1966, to twenty-nine first grade children in one Project school who had been in school for four months and to seventeen preschool children who had been in school two to four days.

---

[1]According to Ott, the number of children tested was seventy-five. Ott, "A Study of Oral English," p. 19.

[2]Jameson, "Phonemic Analysis," p. 18.

[3]Ibid., p. 19.

During administration of the test, it was further revised by dropping the use of the pictures and by omitting one of each of the minimal pair sentences.

The second revision contained fifty-three short test sentences checking twenty-five phonemic problems illustrated with 110 pictures. Each sentence was modeled on a tape and the child was asked to repeat what he heard. There was one word in each sentence that was scored. The score sheet had places for three types of responses to be marked:

(1) the correct word;

(2) the word with a Spanish sound substituted; or

(3) an unpredictable response, a response about which the scorer could not make a decision, or no response.

This test was administered to twenty-one preschool children and twenty-two first grade children in one Project school.

Two teachers, four linguists and two native Spanish speakers studying the teaching of English as a foreign language scored one or more of the tests. All scorers agreed on words heard and marked as correct. There was disagreement, however, on how to score the "incorrect" responses. This indicated need for a third revision.

The third revision contained fifty-seven short sentences covering twenty-five phonemic problems. The sentences were independent of each other in meaning. A voice on the tape said each sentence leaving a pause for a child to

repeat the sentence. The scoring system was the same as for the second revision.

It is interesting that after several more revisions, this test as used by the Southwest Educational Development Laboratory (SEDL) in the 1970-71 school year (under the title Ott Test of Oral Language) had thirty-three sentences covering eleven problem areas. Seventeen of the sentences are exactly the same as the ones in Jameson's third revision. Another one had the word price changed to prize. None of the eleven problems were different from Jameson's.[1]

The SEDL scoring system had three categories for rating a response: correct, minimum change, and no response or "I don't know." The minimum change category is the same as Jameson's category called "the word with the Spanish phoneme substituted." Then, instead of merely reporting the responses of each child or group according to the percentage in each category, the SEDL version assigns four points to a correct response, one point to a minimum change response and zero to the third category. No rationale for this method of scoring was given.

Jameson compared her scoring of the tests with that of thirty-four first, second, and third grade teachers of disadvantaged Spanish-speaking children. As with the second

---

[1]Southwest Educational Development Laboratory, Ott Test of Oral Language, Southwest Educational Development Laboratory, Austin, Texas, 1970.

revision, the various scorers agreed on English phonemes heard as clearly correct, but otherwise individual scoring analyses differed. Despite this, Jameson felt that her third objective was met.

The third revision was administered to the following types of pupils:

(1) Grade One (seven to eight months in school):
twenty OAE pupils in San Antonio;
twenty OAS pupils in San Antonio;
twenty NOA pupils in San Antonio;
twenty Control pupils in San Antonio
(No further description provided);
thirteen Control pupils in Austin
(No further description provided);
eight Anglo-American pupils in Austin
(native English speakers attending a school with an enrollment of more than fifty percent Spanish-speakers);

(2) Preschool (three months in school):
twenty-seven pupils in San Antonio
(No further description provided);
fifteen pupils in Austin
(No further description provided);

(3) Grades Two through Six (sixteen to fifty-two months in school):

fourteen pupils in Austin

(No further description provided).

The major parts of Jameson's dissertation include:

(1) A contrastive outline of English and Spanish which supposedly could be used "as a guide to the teacher without a linguistic background."[1] This outline appears to be somewhat too condensed and overburdened with new vocabulary to be useful to teachers as it stands.

(2) A description of each test item of the third revision of the test, including:

(a) the reason for including the item;

(b) likely deviations from the standard pronunciation;

(c) reasons for the deviations;

(d) data for responses of each group tested, i.e., the percentage of responses in each of the scoring categories;

(e) a summary of the data for each item including, among other things, whether the results were those expected, and relating the results of each item to results of other similar items;

(f) recommendations for changes in certain test items.

---

[1]Jameson, "Phonemic Analysis," p. 26.

This section should be of some value to language program planners, especially the data on the San Antonio OAE, OAS and NOA groups for whom other data are available. Of special interest are the findings that preliterate children do not have trouble with all the same English phonemes with which older children and adults exhibit problems.

(3) A brief summary of the basic difficulties in oral English "discrimination/reproduction" for first graders after: (a) three months of school; and (b) seven or eight months of school. This summary includes comments about the actual errors in comparison to those predicted. For instance, "new vowels in English created much less difficulty than predicted."[1]

Peña

Peña's study is the only one of all the studies concentrating on oral language which focuses on syntax.[2] His dissertation will be referred to as "A Study of Syntactical Structures." His intention was "to ascertain the status of some of the syntactical structures in the oral language development in Spanish and English of four groups"[3] of first

---

[1] Ibid., p. 146.

[2] Albar A. Peña, "A Comparative Study of Syntactical Structures of the Oral Language Status in Spanish and English of Disadvantaged First-Grade Spanish-Speaking Children," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1967.

[3] Ibid., p. 2.

grade children in five Project schools. His study was done during the 1966-67 school year, which was Year Three. At that time some of the first grade experimental pupils were still receiving either special oral language instruction in Spanish or in English, but not in both. These groups were the OAS and OAE groups. Other experimental children were receiving special instruction in science but no structured oral language practice. These children made up the NOA group. The control group was at this time enrolled in the same schools as the experimental children, but received no special science or oral language instruction. Peña referred to this group as the NOA-NS group.

Peña went on to say that his study would focus on "an intensive comparative analysis of some basic sentence patterns and fundamental transformations in Spanish and English manifested in the responses of the subjects at the beginning and at the end of the first grade."[1] A secondary purpose was to "ascertain the usefulness of a new testing instrument purporting to measure basic language development both in Spanish and English."[2]

The title of his dissertation and his stated intentions are somewhat misleading in that the major task Peña performed was to compare the four groups' initial scores and

---

[1] Ibid.

[2] Ibid., p. 3.

their mean gains. These comparisons revealed some information about the effectiveness of the experimental language programs, but said little about the relative or absolute status of the oral language structures of the children.

In order to "ascertain the status" of some syntactic structures and to make a "comparative analysis," Peña collected data as follows:

The pretest instruments, administered in September, 1966, were:

(1) The "Spontaneous Language" section of the Language-Cognition Test;[1]

(2) The Goodenough-Harris Draw-A-Man Test (an intelligence test);[2] and

(3) The Test of General Ability, Inter-American Series (to measure reading readiness), English and Spanish forms.[3]

The posttest instruments, administered in April, 1967, were:

(1) The "Spontaneous Language" section of the Language-Cognition Test; and

---

[1] Anne O. Stemmler, The LCT, Language-Cognition Test, Research Edition (Mimeographed), Austin, Texas, 1967, pp. 1-2.

[2] F.L. Goodenough and Dale B. Harris, Goodenough-Harris Draw-A-Man Test, Harcourt Brace and World, Inc., New York, 1963.

[3] Herschel T. Manuel, Test of General Ability, Inter-American Series, Level 1, Primary, Form CEs, Form DE, Guidance Testing Associates, Austin, Texas, 1962.

(2) The Test of Reading, Inter-American Series, English and Spanish forms.[1]

In the Spontaneous Language section of the Language-Cognition Test, the child had two tasks to perform. For the first task, the child was handed five different familiar concrete objects and asked to name each one and tell everything he could about it. For the second task, he was handed six different pictures and asked to make up a story that would go with the pictures or tell what he saw happening in the pictures. The child's responses were recorded. The test was given twice by the same examiner, once in English and once in Spanish.

Peña could have contributed greatly to the discussion on scoring by presenting at least one example. A lack of clarity occurred because his description in Chapter III, Stemmler's discussion in Appendix A and his further comments in Appendix D are inconsistent. Apparently each child's oral responses were classified and counted. The classification system included twenty-two variables, which were listed only in the appendices.

The first six classifications or variables corresponded to the six general sentence patterns described by

---

[1]Herschel T. Manuel, Test of Reading, Inter-American Series, Primary Level 1, Form ECs, Form DE, Guidance Testing Associates, Austin, Texas, 1962.

Stockwell, Bowen and Martin.[1] These six patterns "allow the framing of an almost endless number of sentences each containing only a single finite verb."[2] This type of description of sentences, however, has certain disadvantages. For instance, "it does not reveal relationships of parts of sentences; it does not show how one kind of sentence may be related to another; and it is not extensible" and it can lead to analyses which are misleading about underlying structures.[3]

The other sixteen classifications or variables included the transformations of negation, interrogatives, affirmative and negative imperatives, passives and subjunctives; one-word utterances; functionally complete sentences; functionally incomplete sentences; Spanish loan words; English loan words; Hispanicized English words; non-standard subject-verb agreement or verb usage; adjectival usage; compound sentences; complex sentences and direct/indirect quotations. Peña provided no explanation for how responses would receive scores in these categories although some are especially vague, such as verb usage and adjectival usage.

---

[1]Robert P. Stockwell, J. Donald Bowen, and John W. Martin, The Grammatical Structures of English and Spanish, The University of Chicago Press, Chicago, 1965.

[2]Ibid., p. 24.

[3]Ibid., p. 4.

After the pretest data were collected, the first task Peña performed was a factor analysis on the frequencies of the various responses from the Language-Cognition Test. The purpose was to reduce the twenty-two variables to a more manageable number of scores, which were then referred to as the LCT factor scores.

Peña then applied analysis of covariance to the LCT factor scores, using the Draw-A-Man Test IQ scores as the covariable. The purpose of this analysis was to find out if there were significant differences between the groups, including sex, on the pretest factor scores.

After the posttest data were collected, Peña performed another factor analysis on the frequencies obtained in the Language-Cognition Test. Then he used repeated measurement analysis of variance to determine if there were significant differences between the factor scores at pre- and posttest for each sex, group and sex-group. Peña's analysis is somewhat limited by his failure to use factor matching techniques or to base his comparisons on raw data rather than factor scores. Factor matching techniques would have indicated the extent to which the factors in the two analyses (pre- and posttest) agreed with one another.

The last thing Peña did was to determine the extent to which correlation existed between the posttest LCT factor scores in English and Spanish and the Test of Reading, Inter-American Series.

Peña apparently never used the scores from the Test of General Ability, Inter-American Series which he collected as pretest data.

Peña chose his subjects randomly from twenty-three first grade classrooms in five of the schools in the Project. All were native Spanish speakers. He originally chose twenty-two pupils in each of the four treatment groups. Due to attrition, he ended with samples of sixteen from each group, eight boys and eight girls. The total N of sixty-four pupils may be too small to produce a reliable estimate of population variance for multivariate (factor analysis) techniques.

Following is a summary of Peña's results: The twenty-two variables of the Language-Cognition Test given as a pretest were reduced to six LCT factors named as follows:

Spanish:
(1) General Sentences and Transformations;
(2) Functionally Complete Sentences;
(3) Basic Sentences with English Loan Words;
(4) Single Words;
(5) Verb Usage; and
(6) Combined Complete and Incomplete Basic Sentences.

English:
(1) General Sentences;
(2) Passive Transformations;

(3) Sentence Fragments;

(4) Functionally Complete Sentences;

(5) Simple Transformations; and

(6) Lack of Negative Transformations.

In his comparison of Goodenough-Harris IQ scores, he found no significant differences between the four treatment groups. In his analysis of covariance using IQ as the covariable, his purpose was to discover if there were significant differences on the LCT factor scores among the groups with the boys and girls separated and between the boys and girls in each group. He did not look for significant differences between groups with the boys and girls together.

For the Spanish factors, analysis of covariance was not justified for Factor 1. He found significant differences among the groups with boys and girls separated for Factor 2 ($p = .001$), Factor 5 ($p = .05$), and Factor 6 ($p = .03$) and between the boys and girls in each group only for Factor 4 ($p = .02$). He attempted to account for these differences by presenting a table of adjusted means for each treatment group divided by sex and by explaining which group(s) caused the major significance.

For the English factors, analysis of covariance was not justified on Factors 1 and 4. He found a significant difference among the groups with boys and girls separated

only on Factor 6 (p = .032) and no significant difference between boys and girls in each group.

All in all, the majority of the findings revealed no significant differences in the pretest scores, including sex, with IQ as the covariable. Peña concluded that:

> In general the groups were found to be relatively similar in their linguistic performance. This conclusion can probably be accounted for by the kind of scoring used.[1]

He stated further that:

> The extremely fundamental nature of the basic sentences and the few transformations employed in the linguistic categories contained the Linguistic Analysis Form did not, apparently permit linguistic distinctions among the groups to be arrived at statistically. . . . Consequently, any future analysis will have to employ more sophisticated linguistic categories (embeddings in particular, for syntax) and a much more in-depth semantic analysis.[2]

The twenty-two variables on the posttest administration of the Language-Cognition Test were also reduced to six factors for both the English and Spanish versions. The LCT factors were named as follows:

Spanish:
(1) General Sentences and Transformations;
(2) Transformations and Complex Sentences;
(3) Single Words and Sentence Fragments;
(4) Basic Sentences and Subjunctive Transformations;
(5) Verb Usage; and

---

[1] Peña, op. cit., p. 109.

[2] Ibid., p. 108.

(6) Simple Sentences.

English: (1) General Sentences;

(2) Complex Sentences;

(3) Complex Sentences and Simple Transformations;

(4) Basic Sentences and Complex Transformations;

(5) Single Words and Sentence Fragments; and

(6) Passive Transformations.

In order to perform repeated measurement analysis of variance on the pre- and posttest factor scores, Peña had to apply the factor weights obtained from the pretest factor analysis to the posttest variables. This was done to make the two sets of factor scores comparable.

Then Peña tested the six English and six Spanish factor scores with three hypotheses:

(1) There are no differences between the LCT factor scores at T_1 (pretest) and T_2 (posttest) for each sex;

(2) There are no differences between the LCT factor scores at T_1 and T_2 for each group; and

(3) There are no differences between the LCT factor scores at T_1 and T_2 for each sex-group combination.[1]

---

[1] Ibid., p. 80.

These mean that Peña was comparing the LCT factor scores of all groups during pre- and posttesting to determine if there were any differences between testing times as a function of treatment and sex.[1]

Peña found that he had to accept all three null hypotheses ($p > .05$) for each factor except:

(1) Spanish Factor 3, second hypothesis ($p = .02$). A comparison of the means of the groups on the factor indicated that the mean increases over time were not the same. The NOA-NS group showed the greatest increase.

(2) Spanish Factor 6, third hypothesis ($p = .03$).

(3) English Factor 5, second hypothesis ($p = .006$). Again the NOA-NS group showed the greatest mean increase.

Peña concluded from these findings that:

> . . . regardless of the treatment used the results obtained will be essentially the same. There would seem to be several explanations for this conclusion. First, since all the groups are being taught with essentially the same program except for the intensive one-hour-per-day instruction in oral language for the OAE and OAS groups, the opportunities for differences in oral language to be expressed by the four treatment groups would probably be minimal during the first year of academic training. Second, it appears possible that the oral language program in its present stage is not intensive enough in its coverage of the level of syntactic maturity of which the child of this age is capable, i.e., simple relative transformations, simple deletions, and so on. Third, insufficient transfer of these types of sentences into other general areas, i.e., fields other than science and the self-concept,

---

[1] Ibid., p. 15.

> which the Spontaneous Language Section of the LCT explored, may have occurred. Finally, it is possible that the syntactical analysis performed may not have had sufficient depth to encompass all the differences manifested in the responses of the subjects being tested.[1]

The last analysis Peña performed consisted of determining the correlations between each of the posttest Spanish LCT factor scores and the Spanish version of the Test of Reading, Inter-American Series and between the posttest English LCT factor scores and the English version of the Test of Reading, Inter-American Series.

Two Spanish factors had significant correlations with the reading test: Factor 1, General Sentences and Transformations (r = .3910) and Factor 2, Transformations and Complex Sentences (r = .2157).

Two English factors had significant correlations with the reading test: Factor 1, General Sentences (r = .4160) and Factor 3, Complex Sentences and Simple Transformations (r = .3789).

Peña concluded that the first section of the Language-Cognition Test:

> . . . did yield evidence of oral language judging by the considerable number of responses obtained in each linguistic category for each subject. . . . The test also proved to have great appeal for children, judging from the enthusiasm that was expressed toward the items and pictures used to elicit "free" responses.[2]

---

[1] Ibid., pp. 109-110.

[2] Ibid., p. 108.

He also recommended that more complicated linguistic analyses be done and that programs of intensive oral language instruction include more sophisticated transformations in the first year and continue more than one hour a day.

Studies on Reading

Arnold

Arnold compared the scores on reading tests administered to the same children in the spring of second grade (May, 1966) and the fall of third grade (September, 1966).[1] His sample consisted of 287 children randomly selected from approximately 700 children who were in Sample I (children who entered first grade in the first year of the Project, 1964-65). The tests used were Metropolitan Achievement Tests, Primary Level II; Test of Reading, Inter-American Series, Level 2; and Prueba de Lectura, Serie Interamericana, Nivel 2.[2] The purpose of making the comparisons was to determine if any significant changes had occurred during the summer months and if the treatments (OAE, OAS and NOA)

---

[1]Richard D. Arnold, "Retention in Reading of Disadvantaged Children," in Reading and Realism, ed. by J. Allen Figurel, International Reading Association, Newark, Delaware, 1969.

[2]Walter N. Durost (ed.), Harold N. Bixler, Gertrude H. Hildreth, Kenneth W. Lund, Metropolitan Achievement Tests, Primary, Level II, Harcourt, Brace and World, New York, 1958; Herschel T. Manuel, Test of Reading, Inter-American Series, Primary, Level 2, Guidance Testing Associates, Austin, Texas, 1962; and Herschel T. Manuel, Prueba de Lectura, Serie Interamericana, Primario Nivel 2, Guidance Testing Associates, Austin, Texas, 1962.

differed with respect to the magnitude or direction of the change. He used the t-test for differences between correlated means to test for significant differences between the spring and fall scores on each test for each treatment group. He also computed difference scores for each pupil and then used the t-test for differences between independent means to compare the means of the difference scores on each test for each pair of treatment groups.

To present his results, Arnold constructed tables showing:

(1) The differences between the spring and fall for each test and treatment. (Each test had three subtests; the scores for each subtest and a total score were included in the analyses.)

(2) Comparison of the NOA and OAS difference scores for each test.

(3) Comparison of the NOA and OAE difference scores for each test.

(4) Comparison of the OAS and OAE difference scores for each test.

The performance of the NOA group declined significantly between the spring and the fall on all three tests. The OAE pupils more or less maintained their position. The OAS group improved its scores significantly on the Test of Reading, Inter-American Series and showed gains on the Metropolitan Achievement Tests.

Arnold concluded that the results appeared to indicate that the oral-aural treatments facilitated retention in reading over the summer. Why the OAS group showed greater gains on an English reading test than the OAE group was an unexplained phenomenon.

Fowler

Fowler's study evaluated the Brengelman-Manning Linguistic Capacity Index as a predictor of reading achievement as measured by the Word Knowledge, Word Discrimination and Reading subtest scores of the Metropolitan Achievement Tests.[1] She tested a random sample of first grade students in eight Project schools during Year Four (1967-68). These pupils were in treatment groups receiving instruction with curricula similar to that of the Sample I and II classes except that the lessons were taught either in English or bilingually (English and Spanish). If the lessons were taught bilingually, either one teacher taught the lessons in both languages or a bilingual teacher and a monolingual (English-speaking) teacher shared responsibilities for two classes. In addition to the original language lessons based on science content, there were also oral language lessons

---

[1]Elaine D. Fowler, "The Brengelman-Manning Linguistic Capacity Index as a Predictor of Reading Achievement of Spanish-Speaking First Grade Children," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1969; Walter N. Durost, ed., Harold H. Bixler, Gertrude H. Hildreth, Kenneth W. Lund, and J. Wayne Wrightstone, Metropolitan Achievement Tests, Primary I Battery, Form C, Harcourt, Brace and World, Inc., New York, 1961.

based on social studies concepts and lessons designed to develop self-concept.

The original sample was a random selection of 152 subjects from thirty-seven classrooms. Due to attrition, the final sample size was ninety-nine students. Fowler did not report the number of classrooms these students represented.

The following data were collected to serve as independent variables in the statistical analyses:

(1) Scores on the Goodenough-Harris Draw-A-Man Test, administered in September, 1967. The Draw-A-Man Test is used to estimate intelligence. It requires no verbal responses from the child.

(2) Scores on the Brengelman-Manning Linguistic Capacity Index, administered in September, 1967. This test measures receptive language ability: Vocabulary, Phonology and Grammar and Total.

(3) Total scores on the Metropolitan Readiness Tests, Form A, administered in January, 1968. The total score is a summation of the scores for Word Meaning, Listening, Matching, Alphabet, Numbers and Copying.

The following data were collected to serve as dependent variables: Word Knowledge, Word Discrimination and Reading scores of the Metropolitan Achievement Tests, administered in April, 1968. In the Word Knowledge subtest the child chooses a word to match a picture. The Word

Discrimination subtest measures the ability to choose the correct written word to match one pronounced by the examiner. The Reading subtest measures comprehension of sentences and paragraphs.

The following null hypotheses were tested:

(1) There will be no significant correlation between the four Linguistic Capacity Index scores and the three Metropolitan Achievement subtest scores.

(2) There will be no significant gain in predictive efficiency when the Linguistic Capacity Index measures are added to information about the subjects' intelligence, as measured by the Goodenough-Harris Draw-A-Man Test, sex and treatment group membership.

(3) There will be no significant gain in predictive efficiency when the Linguistic Capacity Index measures are added to the total scores of the Metropolitan Readiness Tests.

For Hypothesis I, since there were four different Linguistic Capacity Index scores and three Metropolitan Achievement subtest scores, there were twelve different correlations to be computed. The null hypothesis was rejected since all the correlations were significantly different from zero at the .0001 level. They ranged between $r = .46$ for Word Discrimination (Metropolitan Achievement Tests) and Phonology (Linguistic Capacity Index) and $r = .58$ for Reading (Metropolitan Achievement Tests) and the total Linguistic Capacity Index score. These correlations compare favorably

with those generally found between reading readiness and reading achievement.[1]

For Hypothesis II, Fowler presented a table entitled "Summary of Evaluations of the Usefulness of each Brengelman-Manning Linguistic Capacity Index Predictor when Used with Sex, Intelligence, and Treatment Group Membership."[2] This table showed for each Linguistic Capacity Index score predicting each criterion of reading achievement the following: (1) the efficiency of a prediction system containing information about intelligence, sex and treatment group membership and the Linguistic Capacity Index measure; (2) the efficiency of the same system without the Linguistic Capacity Index score; and (3) the predictive usefulness of the Linguistic Capacity Index variable. The F's were all significant beyond the .0005 level indicating that the Linguistic Capacity Index scores were all significant in increasing prediction of Metropolitan Achievement Test scores. Thus the null hypothesis was rejected. The additional variances accounted for by the Linguistic Capacity Index scores ranged from eleven percent for Reading (Metropolitan Achievement Tests) and Phonology (Linguistic Capacity Index) to twenty-three

---

[1]Samuel Weintraub, "Readiness Measures for Predicting Reading Achievement," The Reading Teacher, 20:551 (March, 1967); Robert L. Hillerich, "An Interpretation of Research in Reading Readiness," Elementary English, 43:360 (April, 1966).

[2]Fowler, op. cit., p. 105.

percent for Word Knowledge (Metropolitan Achievement Tests) and Grammar (Linguistic Capacity Index).

To display the results of testing Hypothesis III, Fowler presented a table entitled "Summary of Evaluations of the Usefulness of Each Brengelman-Manning Linguistic Capacity Index Predictor when Used with Reading Readiness Scores as Measured by the Metropolitan Readiness Tests."[1] This table showed for each Linguistic Capacity Index score predicting each criterion of reading achievement the following: (1) the efficiency of a prediction system containing the Metropolitan Readiness Tests total score and the Linguistic Capacity Index measure; (2) the efficiency of the same prediction system without the Linguistic Capacity Index variable; and (3) the predictive usefulness of the Linguistic Capacity Index variable. The F's showed that six of the twelve usefulness values were not significantly different from zero. If .05 is considered the acceptable level of significance, the six values significantly different from zero were:

(1) MAT Reading and LCI Phonology: $p = .03$

(2) MAT Word Knowledge and LCI Grammar: $p = .02$

(3) MAT Reading and LCI Grammar: $p = .004$

(4) MAT Word Knowledge and LCI Total Score: $p = .02$

---

[1] Ibid., p. 108.

(5) MAT Word Discrimination and LCI Total Score: $p = .04$

(6) MAT Reading and LCI Total Score: $p = .004$

Notice that the predictive usefulness of the Linguistic Capacity Index Total Score was significantly greater than zero for predicting each one of the Metropolitan Tests subtest scores. Nevertheless, the null hypothesis could not be rejected in its entirety.

In summary, Fowler found:

(1) Significant correlations between each of the Linguistic Capacity Index scores (Vocabulary, Phonology, Grammar and Total) and three subtests of the Metropolitan Achievement Tests (Word Knowledge, Word Discrimination and Reading).

(2) Significant gain in predictive efficiency when Linguistic Capacity Index subtest and total scores were added to information about the subjects' intelligence (measured by the Goodenough-Harris Draw-A-Man Test), sex and treatment group membership in order to predict reading achievement as measured by the Metropolitan Achievement Tests.

(3) Significant gain for six of twelve possibilities when Linguistic Capacity Index subtest and total scores were added to Metropolitan Readiness Test scores in order to predict reading achievement.

Fowler presented the results of her data analysis in a clear and straightforward manner. Addition of certain items would have made the study somewhat more informative. For instance, before listing the correlations necessary for testing Hypothesis I, she might have given the means and standard deviations on each test for the total sample, for the boys, for the girls and for the members of each treatment group.

Second, since she was asking about the predictive usefulness of the Linguistic Capacity Index in a system which included information about group, sex and IQ, it would have been useful if she had explained in what way group, sex and IQ each contributed to the prediction.

Third, since treatment group membership was an important factor, then the treatments should have been described in greater detail. Moreover, anyone wishing to replicate the study would be unable to do so, given only the information about the experimental treatments in Fowler's study.

Pauck

Pauck's study primarily evaluated Part II (Spanish-English Language Fluency or SELF-Test) of the Ott-Jameson Test as a predictor of reading achievement of Spanish-

speaking first grade children.[1] Reading was measured by the Word Knowledge, Word Discrimination, and Reading subtest scores of the Metropolitan Achievement Tests. He also attempted to determine the predictive effectiveness of the Metropolitan Readiness Tests, Form A, for that population. Finally he compared the effectiveness of the two predictors. His study will be referred to as "Evaluation of SELF-Test."

The sample Pauck tested was identical to the one described by Fowler.[2] The only difference was that Pauck's final sample size was reduced to ninety-seven, two less than Fowler's.

The SELF-Test and the Metropolitan Readiness Tests were administered in January, 1968. The Metropolitan Achievement Tests were administered in April, 1968. The time period between administration of the instruments to be used as predictors and the administration of the achievement test was so short that the findings may be trivial.

Pauck tested five null hypotheses:

(1) The correlation between total word count from the SELF-Test and scores on the reading subtests of the Metropolitan Achievement Tests (MAT) is zero;

---

[1]Fredrick G. Pauck, "An Evaluation of the SELF-Test as a Predictor of Reading Achievement of Spanish-Speaking First Grade Children," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1968.

[2]Fowler, "Evaluation of the Linguistic Capacity Index."

(2) The correlation between the total scores on the Metropolitan Readiness Tests and the reading subtests of the Metropolitan Achievement Tests (MAT) is zero;

(3) The total word count from the SELF-Test is not a significant predictor of reading achievement as measured by the Metropolitan Achievement Tests when sex, treatment and intelligence are statistically controlled;

(4) The Metropolitan Readiness Tests total score is not a significant predictor of reading achievement as measured by the Metropolitan Achievement Tests when sex, treatment and intelligence are statistically controlled; and

(5) There is no significant difference between total word count from the SELF-Test and the Metropolitan Readiness Tests as predictors of reading achievement as measured by the Metropolitan Achievement Tests.

Intelligence was measured by the Goodenough-Harris Draw-A-Man Test, administered in September, 1967. Pauck included a brief description of the oral language programs the children participated in between January and April which were the "treatment" mentioned in the hypotheses. He gave no account, however, of the type of reading instruction being used. He did not say when during the year formal reading instruction began. This lack of information makes it unwise to generalize Pauck's findings to any other population or situation.

The SELF-Test is that part of the Ott-Jameson Test where children are asked a series of questions about what they see in pictures on a filmstrip. The questions were said to represent three levels: literal, non-literal/ inferential and imaginative.

The SELF-Test was contained in a two-track tape cartridge so that the questions are asked on one track and the child's answers are recorded on the other. Scoring was done by counting the words each child said in responding to the questions. As Pauck mentioned, the fact that the child is given only a limited time to answer before another question is asked, means the child may be penalized if he wishes to think about his answer before responding.[1]

Pauck reported that he had originally planned to use two types of scores from the SELF-Test in his data analysis: Word Count, Spanish Intonation and Word Count, English Intonation. He found, however, that the children responded almost exclusively with Spanish intonation.[2] This finding raises questions about Ott's scoring procedures in administering the same test two years before in the same schools. She found a distinction between responses using English intonation and those using Spanish intonation.[3]

---

[1] Ibid., p. 151

[2] Ibid., p. 16.

[3] Ott, "A Study of Oral English."

Pauck's primary method of analysis consisted of:

> . . .correlation techniques and prediction from correlations by means of regression equations, incorporating testing for the significance of the correlation coefficient.[1]

Although he reported in an appendix the raw scores of each child on each test, he did not provide means and standard deviations for the variables being correlated as McNemar suggested is appropriate for correlational studies.[2]

In testing Hypotheses 1 and 2, Pauck found these correlations:

	MAT Word Know.	MAT Word Disc.	MAT Reading
SELF-Test	.627	.597	.657
Metropolitan Readiness Tests	.598	.599	.537

In testing Hypothesis 5 he compared the SELF-Test and the Metropolitan Readiness Tests as predictors and found that the differences between the correlation coefficient of each MAT subtest with the Metropolitan Readiness Tests and the SELF-Test were non-significant. He concluded that the SELF-Test and the Metropolitan Readiness Tests are approximately equally good predictors of reading achievement as measured by the Metropolitan Achievement Tests.

The results of testing Hypotheses 3 and 4 showed that the correlations between the SELF-Test and the MAT

---

[1]Pauck, "Evaluation of SELF-Test," p. 19.

[2]Quinn McNemar, Psychological Statistics, 3rd ed., John Wiley and Sons, Inc., New York, 1962, p. 168.

reading subtests and between the Metropolitan Readiness Tests and the MAT reading subtests are "indicative of a relationship largely independent of sex, treatment, and intelligence."[1]

Pauck felt that, "In light of the above conclusions it would appear that there exists a high degree of association between oral English fluency and reading achievement."[2] This assumes, among other things, that fluency is adequately measured by a word count in a timed test situation.

One possible explanation for the lack of pronounced improvement in the reading abilities of the children in the groups receiving language instruction is that the content of the language materials was not similar to the material in the basal reading programs used in the San Antonio schools. In order to combat this problem, Arnold, in conjunction with his students at The University of Texas at Austin and the Southwest Educational Development Laboratory, developed some beginning reading lessons which had the same content as the oral language materials. The children read no sentence that they had not previously learned to produce in oral language lessons.[3]

---

[1]Pauck, "Evaluation of SELF-Test," p. 150.

[2]Ibid., p. 156.

[3]A more complete description of the experimental reading program can be found in Richard D. Arnold, "English as a Second Language," The Reading Teacher, 21:634-639 (April, 1968).

Melton

Melton used the Word Knowledge, Word Discrimination and Reading subtests of the Metropolitan Achievement Tests, Primary I Battery, Form A, administered in May, 1968, to compare reading achievement of three classes of first grade children who received instruction with the experimental reading materials to the reading achievement of three classes of children who received instruction in reading with the basal reading program.[1] All six classes received instruction in the experimental oral language materials.

Melton's study was done in Year Four (1967-68). By this time the first grade classes in the Project were receiving oral language instruction either in an ESL (English as a Second Language) mode or bilingually. Melton did not say what type of oral language classes her research sample represented. Nor did she have the information to describe the time spent in reading instruction on a daily basis in each classroom. She was not able to report whether or not the teachers using the experimental reading materials used only those lessons or used basal readers in addition. Another unknown variable was the time of year that reading instruction was begun by each teacher.

---

[1]Jo Anne C. Melton, "A Comparison of Two Methods of Teaching Reading to Disadvantaged Mexican-American First Grade Children," Unpublished Master's Thesis, The University of Texas, Austin, 1968.

Melton used one-way analysis of variance to compare:

(1) The mean scores of all the children in the experimental reading classes with those in the basal reading classes;

(2) The mean scores of the classes using the experimental reading lessons; and

(3) The mean scores of the classes using the basal reading program.

On all three subtests and total reading score the basal reading pupils scored higher than the experimental reading pupils. The differences were statistically significant at the .05 level for all comparisons except Word Discrimination. There were significant differences at the .05 level among the three experimental reading classes on Word Discrimination and Reading subtests and total mean scores but not on Word Knowledge. There were significant differences at the .001 level among the three basal reading groups on all three subtests and the total score.

Melton concluded that <u>if</u> a standardized reading test is the criterion for success in reading achievement, "the basal reading program, despite great variation among classrooms, is the better method to teach this group of bilingual Mexican-American children to read."[1] She labeled this a tentative conclusion in light of several limitations of this

---

[1] <u>Ibid</u>., p. 69.

study, among them the small number of classrooms involved and the fact that no pretest measure to assess initial ability of the various classes was used in the analysis.

The fact that the children in the basal reading groups perfcrmed significantly better on the Metropolitan Achievement Tests is not entirely a surprise since the content of the reading tests bears much more similarity to the material in the basal reading program than to that in the experimental reading program. Cook and Arnold collaborated in the development of a reading test with greater content validity.[1] They constructed two forms of a test with a format similar to that of the reading sections of the Metropolitan Achievement Tests.

Cook

In the Word Knowledge subtest of the Cook-Arnold Reading Test, the child marked the one word from four written ones which best matched a line drawing. In the Word Discrimination subtest the child was supposed to mark the one word out of four which the examiner read orally. Each subtest included thirty-five items. According to Cook, the words chosen for both of these tests came from the vocabulary of the experimental reading program. A few other words were used as distractors.

---

[1]Marilyn H. Cook, "The Development of a Reading Test for an Experimental Reading Program for Disadvantaged Children," Unpublished Master's Thesis, The University of Texas, Austin, 1969.

The first section of the reading subtest had thirteen items in which the child chose one of three sentences which best described a line drawing. The second section consisted of eight items in riddle form. The child selected one of three words which best completed the riddle. The sentences approximated those in the experimental reading program.

Examination of copies of both forms of the Cook-Arnold Reading Test included in the appendices of Cook's thesis showed that the tests were not professionally published and were more difficult to read than the usual standardized tests.

The Cook-Arnold Reading Test was administered in May to four classes using the experimental reading programs and four using the basal reading program. Arnold reported that the classes were the same ones tested by Melton.[1] Unfortunately, due to loss of some of the tests, data were not analyzed for two of the basal reading classes.

Cook computed and reported item difficulties and item-scale intercorrelations for each item. One item had a correlation below .10. Ten items had correlations between .10 and .2399 (Cook reported this range to be between .10 and .72 but this must have been a typographical error.) Eighty items had correlations between .24 and .73. Cook

---

[1]Conversation, April 21, 1972.

reported reliability coefficient of over .80 for each section of the test and of .939 for the total test score.

Cook considered the validity and reliability of the Cook-Arnold Reading Test high enough to warrant its use in comparing the reading achievement of the two treatment groups. The mean scores of the children in the experimental reading classes were higher than the mean scores of the basal reading classes for all the subtests and the total, but the difference was significant only for the reading subtest ($p = .020$). This subtest "was essentially a measure of comprehension. Thus, it appears that the prototype experimental group was able to understand more from their reading than was the basal reading group."[1] This is especially encouraging, if the test content was valid for both types of reading programs, in view of the fact that the experimental reading program was not complete and the teachers had to develop many of their own materials.

Cook found significant differences among the mean scores of the classes in the experimental reading program for Word Knowledge ($p = .0019$), Word Discrimination ($p = .0019$), Reading ($p = .0001$) and Total ($p = .0003$). She found significant differences between the two classes in the basal reading program in Word Discrimination ($p = .003$), Reading ($p = .003$) and Total ($p = .009$).

---

[1] Ibid., p. 73.

Socioeconomic Factors

MacMillan

MacMillan's study investigated the relationships between certain socioeconomic factors and school achievement and attendance.[1] He also compared the attendance of three ethnic groups using temperature and precipitation as concomitant variables. His report will be referred to as "Socioeconomic Factors."

The following hypotheses (stated in null form) are those investigated by MacMillan:

(1) The following socioeconomic independent variables are not significant predictors of the child's school achievement:

(a) occupation of father, or mother, if the major wage earner of the family;

(b) size of the family;

(c) family organization, nuclear or atypical;

(d) sex of the child;

(e) preschool experience; and

(f) the child's number of days of attendance in school;

(2) The following socioeconomic independent variables are not significant predictors of the child's school attendance:

---

[1]Robert W. MacMillan, "A Study of the Effects of Socioeconomic Factors on the School Achievement of Spanish-Speaking School Beginners," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1966.

(a) occupation of the father, or mother, if the major wage earner of the family;

(b) size of the family;

(c) family organization, nuclear or atypical; and

(d) sex of the child;

(3) The following factors are not significant predictors of the child's school attendance:

(a) membership in the Mexican-American ethnic group;

(b) membership in the Negro ethnic group;

(c) membership in the Anglo ethnic group;

(d) membership in upper, middle, or lower social group;

(e) amount of daily precipitation; and

(f) daily temperature.[1]

For the first two hypotheses, the sample consisted of 305 first grade pupils in the Project for whom there were available complete test data and complete socioeconomic data. The study was done during Year One (1964-65). For the third hypothesis, the sample was drawn from five Mexican-American schools, four in the Project and one random; four Negro schools and three Anglo schools. Four of the Mexican-American schools and all the Negro schools were categorized

---

[1] Ibid., pp. 14-17.

by the San Antonio Independent School District as being in the low socioeconomic class. The three Anglo schools were in the upper-, middle-, and lower-class categories.

Socioeconomic data were collected from pupil questionnaires constructed by the Project staff and filled out by the classroom teachers. Attendance data were collected from school registers and teacher reports. For the third hypothesis, the days counted for attendance purposes were the same 140 days as those in which the experimental lessons were taught: September 21, 1964 to April 27, 1965.

Test data were collected from the Metropolitan Readiness Tests, Form A; Prueba de Habilidad General, Serie Interamericana, Primario, Nivel 1; and the Goodenough-Harris Draw-A-Man Test. For Hypothesis 1 the achievement measure was obtained by converting the posttest Metropolitan Readiness Tests scores and the pretest Prueba de Habilidad General scores to standard scores and subtracting pretest and posttest. The rationale for this procedure was not given. The primary method of analysis was multiple linear regression.

MacMillan found that:

(1) The combination of socioeconomic variables of parent's occupation, family size and organization, pupil sex, preschool experience and attendance are predictors of school achievement;

(2) IQ scores and pretest scores were significantly correlated with achievement;

(3) Parent occupation and pupil attendance were equally important predictors of achievement;

(4) Preschool experience did not quite reach statistical significance ($p = .0579$) as a predictor, although it should be considered educationally important;

(5) Family size and organization were not predictors of achievement;

(6) Sex as a single factor was not a significant predictor of achievement;[1]

(7) Individually, none of the socioeconomic variables were important predictors of attendance;[2]

(8) Significant differences in attendance patterns were found among the three socioeconomic groups in the Anglo ethnic group;[3]

(9) Among the Mexican-American ethnic group the low socioeconomic group attended school more frequently than the middle group regardless of weather conditions. He speculated that this reflected the drawing power of the Project;[4]

(10) Attendance of the Negro ethnic group was most affected by weather conditions;[5]

---

[1] Ibid., p. 222.

[2] Ibid., p. 223.

[3] Ibid., p. 224.

[4] Ibid.

[5] Ibid., p. 225.

(11) If attendance is an indicator of attitude, the Mexican-American first grade child has as positive an attitude as does the Anglo child.[1]

MacMillan also provided a review of the demographic and sociological literature pertaining to the Mexican-American of the Southwest. In addition, a demographic description of Mexican-Americans in San Antonio included statistics, pictures and other information on population, income, housing and education. These two sections established very well the context in which the San Antonio Language Research Project operated.

McDowell

McDowell compared the intelligence, general ability and reading achievement of first grade children from three ethnic groups--Negro, Anglo and Spanish-surname--during the first year of the longitudinal study, 1964-65.[2] Intelligence was measured by the Goodenough-Harris Draw-A-Man Test administered in September. General ability was measured by the Test of General Ability, Inter-American Series, administered in September. The Test of Reading, Inter-American Series, administered in May, was used to measure reading ability.

---

[1] Ibid., p. 228.

[2] Neil A. McDowell, "A Study of the Academic Capabilities and Achievements of Three Ethnic Groups: Anglo, Negro, and Spanish-Surname, in San Antonio, Texas," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1966.

The children tested were in twenty-eight classes in nine "experimental" or Project schools and in twenty-one classes in twelve "control" schools. The control schools were selected to represent a cross-section of the San Antonio Independent School District. They included Negro schools, predominantly Spanish-surname schools, upper-class Anglo schools and mixed schools including children from lower-middle class Anglo and Spanish-surname homes. Estimate of social class levels was provided by personnel of the school district.

The Draw-A-Man Test presumably was administered to all children in English. The Test of General Ability, Inter-American Series was administered in Spanish to children in the experimental schools and in English to children in control schools. The Test of Reading, Inter-American Series, was administered in English and Spanish to children in experimental schools and in English only to children in control schools.

McDowell designed his study to test the following hypotheses:

(1) There will be no difference in the pretest (general ability) scores among the Anglo, Negro, and Spanish-surname children in the control schools of Horn's study.

(2) There will be no difference in the posttest (reading achievement) scores among the Anglo, Negro, and Spanish-surname children in the control schools when the related factors of age, intelligence, pretest score, and school attendance are held constant.

(3) The children in the control group who are repeating first grade will not demonstrate significant gains in reading achievement over matched groups of nonrepeaters from the control and experimental schools.

(4) There will be no significant difference among the posttest scores (reading achievement) of Negro children, Anglo children, and the Spanish-surname children of the control schools and the children in the experimental schools when the related factors of age, intelligence, pretest score, and school attendance are held constant.[1]

He also measured the:

> . . . differences between pretest (general ability) and posttest (reading achievement) scores when the children of the control and experimental schools were grouped according to socioeconomic level.[2]

In testing the first hypothesis, McDowell found that the Anglo children scored higher than either the Negro or the Spanish children in all areas of the pretest scores. Significant differences occurred between the scores of the Spanish-surname children and the Negro children on four variables. The differences on the oral-vocabulary subtest of the Test of General Ability, Inter-American Series favored the Negro children. The differences on the association subtest of the Test of General Ability, Inter-American Series and on both the raw scores and standard scores from the Draw-A-Man Test favored the Spanish-surname children.

---

[1] Ibid., pp. 8-9.

[2] Ibid., p. 9.

Because of interaction among the groups on three of the covariables to be considered in testing the second hypothesis, no statement could be made "concerning significance of differences among groups for the continuous range of scores on the criterion variable."[1] This same difficulty also prevented meaningful comparisons from being made as planned in the fourth hypothesis above.

In testing the third hypothesis, Spanish-surname and Anglo repeaters in the control schools and nonrepeaters in the experimental and control schools were matched by standard IQ scores. Scores of three groups were then compared: repeaters in control schools, nonrepeaters in control schools and nonrepeaters in experimental schools. When the total group was the sample, the repeaters scored significantly higher (.05) than the nonrepeaters on two of the subtests on the pretest, Test of General Ability, Inter-American Series. On the posttest, Test of Reading, Inter-American Series, no group scored significantly higher than any other.

When the scores of Spanish-surname children in the control schools were analyzed separately, the repeaters were found to have scored significantly higher ($p = .001$) on the pretest but there were no significant differences at the .05 level on the posttest. When the scores of the Anglo children in the control schools were analyzed separately, the

---

[1] Ibid., p. 103.

differences between the repeaters and nonrepeaters were not significant on either the pretest or posttest.

McDowell also grouped the children according to socioeconomic class and compared their scores on two subtests (oral-numerical and association-classification) of the pretest, the Test of General Ability, Inter-American Series, and on total comprehension-vocabulary of the posttest, the Test of Reading, Inter-American Series. The control classes were designated as upper-class, middle-class, middle-lower class, or lower-lower class. The pupils in the experimental schools, all considered to be lower-lower class, were kept separate for this analysis. For each of the three tests, McDowell presented a table showing the mean scores, the difference between the means, the variances and the difference between the variances for each social class. The rank order of the groups according to the mean scores for both pretests was: middle-lower class, upper-class, lower-lower class, middle-class and experimental group. On the posttest the rank order remained the same except that the middle-lower class dropped from first to fourth place. McDowell did discover, however, in looking at the relationships between each pretest variable and the posttest variable, that the scoring pattern of the middle-lower class was different from that of the other four groups. As it turned out, all the children representing the middle-lower class were Negroes from the same school.

The comparison of the scores of the various social classes revealed several points of interest according to McDowell:

> (1) The children in the experimental classes scored significantly lower as a group on both pretest and posttest than the children in any of the socioeconomic groups of the control schools.[1]
>
> (2) The children in the upper class appear to achieve greater gains in reading than children of the other groups.[2]
>
> (3) It is possible to conclude that there are definite differences among the socioeconomic groups in their scores on the different sections of the two tests in question.[3]

McDowell performed a series of analyses to supplement those which proved inappropriate for testing Hypotheses 2 and 4 due to interaction among the groups. First, he planned to use analysis of covariance to compare the scores of the children in the three ethnic groups in the control schools except for the Negro children representing the middle-lower class. The pretest, the Test of General Ability, Inter-American Series, was the independent variable and the Test of Reading, Inter-American Series was the criterion variable. Age, intelligence and attendance were held constant. This time no significant interaction among the groups was found so analysis of covariance was appropriate. No significant differences

---

[1] Ibid., p. 127.

[2] Ibid., pp. 127-28.

[3] Ibid., p. 128.

were found on the criterion scores among the three ethnic groups in the control schools.

Second, he repeated the first analysis except he also dropped out the scores of the Spanish-surname repeaters. Again he found no significant differences on the criterion scores among the three ethnic groups. Because of the results of the first two analyses, he conjectured that:

> . . . significant differences in group scores are occasioned by an unequal distribution of students at the different ends of the continuous scale of values (whatever the values: e.g., age, intelligence, pretest).[1]

Third, he intended to compare the scores of the children in the control classes, minus the middle-lower class children, with the scores of the children in the experimental schools. Significant interaction prevented this analysis from being completed.

Last, he was going to repeat the third analysis except that the scores of the Spanish-surname repeaters would be excluded. However, again significant interaction found among the groups prohibited completion of the analysis.

In conclusion, McDowell noted that although differences were found among the ethnic groups in the control schools in the pretest and the posttest scores, limitations of the study prevented anything to be said about the

---

[1] Ibid., p. 156.

differences except that "they do exist and the effect on schooling of the groups is not consistent."[1]

Additional Studies

Studies by Bradley, Bussey and Cornejo were also done as part of the San Antonio Language Research Project. All related to oral language of children.[2] The studies by Bradley and Bussey involved more extensive use of the Gloria and David sentence repetition test first used by Taylor.[3]

---

[1]Ibid., p. 148.

[2]N. Ruth Bradley, "A Study of the Relation of Oral Language Proficiency and Reading in a Group of Fourth Grade Children of a French Linguistic Background," Unpublished Master's Thesis, The University of Texas, Austin, 1970; N. Ruth Bradley, "A Two-Year Investigation of the Relation of Oral Language Proficiency and Reading Achievement of First Grade Children with a French Linguistic Background," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1971; Jo Ann Keslar Bussey, "A Comparative Study of Phonological Variations Among First, Second, and Third Grade Linguistically Different Children in Five San Antonio (Texas) Schools," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1971; Ricardo Jesus Cornejo, "Bilingualism: A Study of the Lexicon of the Five-Year-Old Spanish-Speaking Children of Texas," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1969.

[3]Taylor, "Year Five Findings."

SUMMARY

Beginning in the fall of the 1964 a group of Mexican-American "disadvantaged" children in a research project in San Antonio, Texas, received instruction in one of three treatments: special English language instruction using oral-aural techniques with science content, special Spanish language instruction using oral-aural techniques with science content, and instruction with the same science content but with no special language instruction. These treatments were known respectively as the OAE (oral-aural-English), OAS (oral-aural-Spanish) and NOA (non-oral-aural). After three years the NOA treatment was discontinued and the labels of the first two were changed to LCE (language-cognition English) and LCS (language-cognition Spanish). Other changes were made in the treatments which affected only the children who entered first grade in 1966 or later.

The experimental treatments began when the children were in first grade and were continued as they proceeded through elementary school. Most of the research efforts of the staff and students at The University of Texas at Austin had their focus on the effects of the treatments on the children who entered first grade in 1964 and 1965. The research sample consisting of the children who entered school in 1964 was known as Sample I. The children who entered school in 1965 made up Sample II.

Horn found that after one year of treatment there were no significant differences among the scores of the OAE, OAS and NOA groups in Sample I in reading readiness.

Arnold studied Sample I after two years of treatment and Sample II after one year. He compared the test scores of the three treatment groups and a control group selected from a cross-section of San Antonio Schools. The scores were taken from several tests of language, group intelligence, English reading and Spanish reading. In Sample I Arnold found that generally there were no significant differences between the OAE and OAS groups, that the NOA group scores were significantly higher than the OAE and OAS groups and that the control group scores were highest. Although the results for Sample II were inconsistent, it was possible to generalize that the OAE treatment was superior in reading achievement and the control group was superior in reading readiness.

Knight studied Sample I after three years of treatment and Sample II after two years. His analysis included no control group. He compared English and Spanish reading scores using analysis of covariance with pretest scores and group intelligence test scores as the covariables. The results of the comparisons were not consistent but the findings for Sample I tended to favor the NOA group and for Sample II the OAE group.

The research concerning the results after the fourth year is reported in Chapter III, which follows. For the Year Five study, Taylor constructed an oral language test which was intended to measure phonology, intonation and fluency in English. She used the test to assess the language of children who had been in LCE or LCS classes for five years (Sample I) and four years (Sample II) and of children who had never received special language instruction. For Sample I the LCE group received the highest mean score in Phonology. The LCS group received the highest mean scores in Fluency and Total Language. The control group received the lowest mean scores. For Sample II none of the differences was significant.

Ott compared the English language abilities of Sample II children in the OAE and NOA groups when they were in first grade. She used the Brengelman-Manning Linguistic Capacity Index to measure receptive language, i.e., ability to understand oral language; she constructed her own test to measure expressive language, i.e., ability to speak. She concluded that both groups made significant gains in receptive and expressive language. The gains of the OAE group were significantly greater than those of the NOA group in expressive language but there was no significant difference between the groups in amount of gain in receptive language.

Jameson constructed a test for assessing children's ability to produce certain phonemes. She administered the test to some children in Sample II while they were in first grade and to other Spanish and non-Spanish-speaking children from preschool through sixth grade. Jameson's contribution consisted mainly of a contrastive outline of English and Spanish phonology, the instrument for assessing language and a summary of the data collected when the test was administered. She attempted no evaluation of the experimental treatments.

Peña attempted to evaluate control of syntax in the oral English and Spanish of first graders. The instrument he used was the research edition of a test developed by Stemmler. The children tested entered first grade a year later than the children in Sample II. The experimental treatments, however, were approximately the same as in previous years: OAE, OAS and NOA. A fourth group (NOA-NS) consisted of children in the Project schools who had not received special language or science instruction. Peña administered the test in September as a pretest and in April as a posttest. He found in general that there were few significant differences among the four groups on either the pretest or the posttest.

Arnold tested reading achievement of children in Sample I at the end of second grade and the beginning of third grade. He found that over the summer the NOA group

scores declined significantly, the OAE group more or less maintained its position and the OAS group improved its scores.

Fowler administered the Brengelman-Manning Linquistic Capacity Index, the reading tests in the Metropolitan Achievement Tests, and the Metropolitan Readiness Tests to a sample of first graders who entered school in 1967. These children received special oral language instruction either in English or in both English and Spanish. Fowler found significant correlations between each of the Linquistic Capacity Index scores and the three subtests of the Metropolitan Achievement Tests. She also found gains in predictive efficiency when Linquistic Capacity Index scores were added to information about IQ, sex and treatment in order to predict reading achievement as measured by the Metropolitan Achievement Tests. Finally, she found significant gains in predictive efficiency in six of twelve possibilities when Linquistic Capacity Index scores were added to Metropolitan Readiness Tests scores to predict reading achievement.

Pauck analyzed scores of the same children as in Fowler's sample except that he used the fluency section of the oral language test Ott designed instead of the Linquistic Capacity Index. He counted words in the children's responses to certain questions to obtain scores for the oral language test. He found that the oral language test and the Metropolitan Readiness Tests were equally good predictors of

reading achievement as measured by the Metropolitan Achievement Tests.

Melton administered the reading sections of the Metropolitan Achievement Tests to six classes of first grade children receiving the experimental treatments who entered school two years later than the children in Sample II. Three of the classes received reading instruction in a basal reader program and the other three classes used experimental reading materials based on the special oral language treatments. She found that the basal reading pupils scored significantly higher than the experimental reading pupils.

Cook and Arnold constructed a reading test with the same format as the Metropolitan Achievement Tests, but based on the content of the experimental reading program. Cook administered the tests to eight first grade classes including the same ones in Melton's research sample. On this test the experimental reading pupils scored higher than the basal reading pupils but the difference was significant on only one subtest.

MacMillan investigated the relationships between certain socioeconomic factors and school achievement and attendance of pupils in Sample I during their first grade year. He also collected data on first grade pupils of other ethnic groups and socioeconomic classes in the San Antonio school district. He found, among other things, that a combination of socioeconomic variables was a predictor of achievement and

sex as a single factor was not a significant predictor. He discovered significant differences in attendance patterns in the three socioeconomic classes in the Anglo ethnic group. He also found that Negro group attendance was most affected by weather conditions.

McDowell compared the intelligence, general ability and reading achievement of first grade children from three ethnic groups and several socioeconomic classes. For some of his analyses he included children from Sample I. McDowell found significant differences on pretest and posttest scores among the three ethnic groups in the control schools. He also found that the children in the experimental classes received significantly lower scores than those of children in any socioeconomic group in the control schools.

The studies reviewed here had several methodological weaknesses in common. For instance, none of them included adequate descriptions of the experimental treatments. However, since the studies were related, sample lesson plans were only provided in the appendices of one or two which gave a clear idea of the special curriculum. Information such as the nature of the supervision and inservice training was not described in detail. In addition, none of the authors gave a candid explanation of the procedures used in developing the curriculum materials for any of the grade levels.

None of the writers described the regular San Antonio curriculum. Thus, any comparisons of the experimental treatments with the regular curriculum cannot be interpreted meaningfully. Uncertainty as to the time teachers actually devoted daily to the experimental treatments also inhibits accurate interpretation of results.

Several of the studies displayed an unsophisticated use and understanding of statistics. Dissertations involving correlation did not clearly indicate that the authors understood the difficulties inherent in such studies. Also, some remarks reflected confusion with regard to the distinction between correlations high enough to be useful and correlations high enough to be statistically different from zero. In several studies, readers are handicapped by incomplete presentations of data. For instance, in one correlation study and in one study involving analysis of covariance, the means and standard deviations for each group on each variable were not reported. These weaknesses in statistical matters may be at least part of the cause for another disappointing feature of many of the studies. The discussion sections found in the concluding chapters were frequently quite brief. The reader is given too little help in interpreting the data.

Several weaknesses in research design affect the conclusions that can be drawn from the results of several studies. An unknown number of subjects in the Year Two

(1965-66) and Year Three (1966-67) studies did not receive the same treatment for the entire two or three years they were in the Project. The attrition rate was so great that the Year Four (1967-68) study, which is contained in Chapters III and IV of this dissertation, and the Year Five (1968-69) study both suffer from very small sample sizes compared to the original number of pupils in each treatment. The results of a number of still other of the studies would have been more revealing if appropriate control groups had been included in the research design.

Evaluation efforts throughout the life of the Project were handicapped by a lack of tests for assessing various aspects of oral language in English and Spanish. This lack was, of course, the reason why several dissertations consisted of attempts to develop appropriate tests. The authors of these oral language instruments could have made even greater contributions, however, if they had provided complete rationales for the methods they used in constructing the tests. The authors did not establish firm connections between the aspect of language supposedly being measured and the measure itself. In other words, the validity of the tests is questionable. The methods of evaluating responses were not always explained clearly. In addition, the authors did not furnish rationales for the scoring systems they used. A few of the authors of the oral language tests did not describe the conditions under which their tests were administered, the

procedures used in the testing and the procedures for scoring the responses.

The deficiencies in research methodology probably reflect two major types of problems: (1) those associated with attempting longitudinal research in the public schools without strong commitment and sustained cooperation from the school personnel in a project conducted by a small staff supported by a limited budget; and (2) those to be expected in research carried out by individuals who lack extensive training in research methodology and statistical techniques.

CHAPTER III

DESCRIPTION OF PROCEDURES AND RESULTS OF DATA ANALYSES

This study was designed to evaluate the effectiveness of various experimental treatments of the San Antonio Language Research Project at the end of Year Four (1967-68). Scores from the Iowa Tests of Basic Skills and The Lorge-Thorndike Intelligence Tests were analyzed for third and fourth grade children who had received:

(1) Intensive oral-aural instruction in English, with science content. This group will be referred to as the LCE (Language-Cognition English) group.

(2) Intensive oral-aural instruction in Spanish, with science content. This group will be referred to as the LCS (Language-Cognition Spanish) group.

(3) Instruction in science content but without oral-aural language teaching techniques for two years and intensive oral-aural instruction in English for one year. This group will be referred to as the NOA-LCE group.

(4) Instruction in the usual San Antonio curriculum with no experimental curriculum or special inservice activities. This "control" group will be referred to as the SAC group.

The remainder of this chapter relates in detail the operational and statistical procedures used and the results of the data analyses.

DESCRIPTION OF PROCEDURES

Population

Information in Chapter I established the economic and educational plight of Mexican-Americans. Dissertations by MacMillan and Bussey provide information describing the pupils who attend Project schools and their families.[1] In summary, almost all of the children in the Project schools are Spanish-speaking Mexican-Americans of low socio-economic status.

Sampling Procedure

Despite the fact that between six and eight hundred pupils were in experimental classes in each grade in each year of the Project, the sample sizes for research purposes have been smaller than 200 for each treatment. Complete data for every child was difficult to secure, partly because

---

[1]Robert W. MacMillan, "A Study of the Effect of Socioeconomic Factors on the School Achievement of Spanish-Speaking School Beginners," The University of Texas, Austin, 1966; Jo Ann Keslar Bussey, "A Comparative Study of Phonological Variations Among First, Second, and Third Grade Linguistically Different Children in Five San Antonio (Texas) Schools," The University of Texas, Austin, 1971.

of transfers and absenteeism of pupils and partly because of teacher attrition. Table 5 shows the number of pupils in experimental classes during the first three years and the number of subjects in the data analyses.

By Year Four (1967-68), it became apparent that a large number of children were not being promoted with their classmates or were being assigned to non-experimental classes or to classes receiving different treatments from the one they received in first grade. The intention of this study was to evaluate the long-run effectiveness of the experimental treatments. Therefore, it was deemed appropriate to include in the sample only pupils who had received the same type of instruction for their entire school career and who had not failed a grade. (The number who failed was very small.) Naturally, limiting the sample in this way would be one more factor reducing the sample size.

In order to identify the children who had always been in the same treatment, three resources were used. A list of all the pupils for whom scores of The Lorge-Thorndike Intelligence Tests and the Iowa Tests of Basic Skills were available was compiled and compared first to a list supplied to Taylor[1] by Ellen Snow, who was on the staff

---

[1]Thomasine H. Taylor, "The Effects of Oral-Aural Language Training on Gains in English Language for Fourth and Fifth Grade Disadvantaged Mexican-American Children," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1969.

Table 5

Number of Children in Experimental Classes and in Data Analyses Year One (1964-65), Year Two (1965-66) and Year Three (1966-67) San Antonio Language Research Project[1]

Treatment	Grade	Number of Pupils in Exper. Classes	Number of Pupils in Data Analyses
Sample I (Entered first grade 1964)			
LCE	1	250	186
	2	300	178
	3	200	120-160[2]
LCS	1	250	204
	2	250	178
	3	190	114-152[2]
NOA	1	250	194
	2	275	177
	3	250	150-200[2]
Sample II (Entered first grade 1965)			
LCE	1	300	187
	2	304	182-243[2]
LCS	1	175	105
	2	202	121-161[2]
NOA	1	275	160
	2	255	153-204[2]

[1]Compiled from Thomas D. Horn, "A Study of the Effects of Intensive Oral-Aural English Language Instruction, Oral-Aural Spanish Language Instruction and Non-Oral-Aural Instruction on Reading Readiness in Grade One," The University of Texas, Austin, 1966, p. 18; Richard D. Arnold, "Capsule Description of San Antonio Language-Bilingual Research Project," The University of Texas, Austin, 1967 (Mimeographed); Richard D. Arnold, "1965-66 (Year Two) Findings, San Antonio Language Research Project, Thomas D. Horn, Director," The University of Texas, Austin,

of the San Antonio Bilingual Research and Dissemination Center. This list contained the names of all children who had received consistent treatment and had not failed a grade through Year Five (1968-69). This left unchecked the students who received consistent treatment only through Year Four (1967-68). Therefore, two other resources were used: treatment histories prepared by the teachers of experimental classes in Year Four (1967-68) and the Permanent Cumulative Records for each pupil. This process resulted in the sample sizes shown on Table 6.

The children who served as controls were pupils in Project schools who had never been in experimental classes and whose scores were made available to this researcher. The controls received the regular San Antonio curriculum so they will be referred to as the SAC group.

Examination of the ages of the children in the SAC group showed that at least one of the twenty-one fourth grade pupils was overage for his grade and at least eight of the sixty-eight third grade pupils were overage. Thus,

---

1968, pp. 6 and 8; and Lester N. Knight, "1966-67 (Year Three) Findings: A Comparison of the Effectiveness of Intensive Oral-Aural English Instruction, Intensive Oral-Aural Spanish Instruction, and Non-Oral-Aural Instruction on the Oral Language and Reading Achievement of Spanish-Speaking Second and Third Graders," The University of Texas, Austin, 1969, pp. 106 and 127.

[2]Estimates derived from Knight's statement that sixty to eighty percent of the scores survived his matching technique. Knight, op. cit., p. 127.

Table 6

Number of Children in Research Sample
Year Four (1967-68), San Antonio

Treatment Group[1]	Grade Four	Grade Three
LCE	29	109
NOA-LCE	15[a]	59
LCS	60	40
NOA-LCS	17[a]	1[a]
SAC	21	68

[1]LCE = Language-Cognition English; NOA-LCE = Non-Oral-Aural until Year Four and then Language-Cognition English; LCS = Language-Cognition Spanish; NOA-LCS = Non-Oral-Aural until Year Four and then Language-Cognition Spanish; and SAC = San Antonio Curriculum (controls).

[a]Data for these children were not analyzed.

it is possible that some of the SAC pupils had failed at least one grade whereas scores for the few children who had been failed in the experimental groups were eliminated from the data analyses.

Scores in three cells on Table 6 were not used in the data analysis. The cell with one student, Grade Three NOA-LCS, is obviously inappropriate for further analysis. The cells with fifteen and seventeen pupils, Grade Four NOA-LCE and NOA-LCS, were dropped because of small size and because all the children in each cell had the same teacher in Year Four (1967-68). Moreover, the seventeen children in the NOA-LCS group had a first-year teacher.

In Year Four (1967-68) the students in each treatment came from several classes in several schools. This fact may help to moderate the teacher variable effect. Table 7 reveals the breakdown by school and teacher.

Instrumentation

The Iowa Tests of Basic Skills were chosen as the major instrument for the study since they are intended to measure generalized intellectual skills and abilities in third to ninth graders. They are not intended to measure the "acquisition of specific information in special subjects."[1]

---

[1]E.F. Lindquist and A.N. Hieronymus, Manual for Administrators, Supervisors, and Counselors: Iowa Tests of Basic Skills, Houghton Mifflin Co., Boston, 1964, p. 4, hereafter referred to as Manual for Administrators.

Table 7

Number of Teachers and Schools in Research Sample, Year Four (1967-68), San Antonio

School	Number of Teachers		
Grade Four	LCE[1]	LCS[2]	SAC[3]
Crockett		2	
Carvajal	1	1	
Johnson		1	
Ogden	1		3
Ruiz			2
Brackenridge	1	2	
Grade Three			
Crockett	1	2[a]	3
Carvajal	2	1	
Johnson	2	1	
Ogden	2		2
Ruiz	1		2
Brackenridge	4	1	
Brewer	1		

[1]LCE = Language-Cognition English

[2]LCS = Language-Cognition Spanish

[3]SAC = San Antonio Curriculum

[a]These two teachers taught one class, one each semester.

The question remained as to what type of statistical analysis would be most appropriate and useful.

Scores for the Iowa Tests of Basic Skills were available from both fall and spring administrations. Therefore, one possible analysis would have been to compare the mean gain scores of the different groups in order to see if the difference between the fall and spring scores is any greater for one group than any of the others. This alternative was eliminated since the main interest was in what had happened to the children as a result of four or three years of experimental treatment rather than in what happened during Year Four (1967-68) only.

Another possibility would have been to compare the spring scores using analysis of covariance with the fall scores as the covariable. This means that the spring scores would be compared only after the groups had been equated statistically in terms of fall scores. This choice had at least two disadvantages. One, again this would result in looking only at what happened during one school year. Two, a child's ability or achievement level in the fall is due to many factors, among them: IQ, past teachers, other learning experiences in and out of school, and his own diligence. It would be hard to say in exactly what terms the groups were being equated.

Many of the children's test scores from previous years in the Project were theoretically available so that

these scores might be considered covariables in a comparison of the spring scores of the Iowa Tests of Basic Skills. The actual degree of availability was very low, however. Punch cards containing test data from Year One (1964-65) existed but the location of the decoding sheet at the time this study was begun was unknown. All the tests from Years Two (1965-66) and Three (1966-67) were graded and in storage. Unfortunately, disorder in the files, along with the low probability that there would be complete data on any one child, discouraged anyone from investing the great amount of time necessary to do anything with the test scores accumulated during the length of the Project. These difficulties ruled out any analysis using previous years' scores as covariables as well as any analysis of trends in test scores of the groups over the years.

With the previously mentioned alternatives eliminated, the only reasonable analysis left seemed to be to compare the mean scores on the spring administration of the Iowa Tests of Basic Skills. Obviously, a one way analysis of variance could be performed. This involves "testing of the significance of the over-all variation of the means for several groups, the groups differing on the basis of a single classificatory principle."[1] In this

---

[1]Quinn McNemar, Psychological Statistics, 3rd ed., John Wiley and Sons, Inc., New York, 1962, p. 288.

case, the single classificatory principle is the treatment group: intensive instruction in English or Spanish or regular San Antonio curriculum. In all other respects the groups would be assumed to be equal.

Thus, in order to obtain any meaningful conclusions from analysis of variance, one must know whether or not the experimental groups were similar at the beginning of the study. In the case of the San Antonio Language Research Project, no one knows for certain whether there were significant differences in educational factors between the groups at the beginning of Year One (1964-65). For the sake of argument one has to assume that there were no significant differences, although in several experimental classes the youngest children were found to have been deliberately assigned rather than being randomly selected. Analysis of Year One data indicated that the standardized instruments available were not appropriate for measuring whatever differences might have existed in reading readiness and general ability in disadvantaged Spanish-speaking first grade children.

The children throughout the San Antonio Independent School District had administered to them a group IQ test, The Lorge-Thorndike Intelligence Tests, in the fall of third grade. Thus, by Year Four (1967-68) IQ scores for both samples (third graders and fourth graders) were available. How valid these scores are is questionable, but it

is probably safe to say that they are more valid than scores of any test administered at the beginning of first grade. In addition, whatever factors lead to lack of validity probably affect the scores of all the children in this research sample in about the same way. It was decided to perform also an analysis of covariance using IQ as a covariable. Analysis of covariance "is applicable whenever it seems desirable to correct a difference on a dependent variable for a known variable on another variable which for some reason could not be controlled by matching or by random sampling procedures."[1]

In summary, the question to be answered in this study was: Will there be significant differences among the mean scores of the treatment groups on the posttests? The posttest was the Iowa Tests of Basic Skills administered in the spring of 1968. Both analysis of variance and analysis of covariance were to be performed. The covariable was to be IQ measured by The Lorge-Thorndike Intelligence Tests administered in the fall of 1967 for third grade children and in the fall of 1966 for fourth grade children.

The Lorge-Thorndike Intelligence Tests[2]

This is a group intelligence test. The scores used

---

[1]McNemar, op. cit., p. 363.

[2]Irving Lorge and Robert L. Thorndike, Houghton Mifflin Co., Boston, 1957.

in the Year Four (1967-68) analysis are from Level 2, Form A, designed for grades two and three and consisting only of a non-verbal battery. This test is extensively reviewed in The Fifth Mental Measurements Yearbook.[1] For that reason, comments here will be brief.

According to Freeman, the Lorge-Thorndike series are on the whole among the sounder group instruments available.[2] Pidgeon claimed that although one may doubt the value of the non-verbal tests for predicting ability to read or any scholastic performance, the test does adequately serve the purpose of separating children into relatively homogeneous ability groups.[3] He warned that no assumptions should be made about the test measuring mental capacity, but "the tests provide reliable measures of verbal reasoning and nonverbal reasoning."[4]

In this study the assumption is not made that the Lorge-Thorndike tests measure or predict innate ability to learn. Probably the test scores reflect differences in cultural background, in what has already been learned, in

---

[1]Oscar K. Buros, ed., The Fifth Mental Measurements Yearbook, The Gryphon Press, Highland Park, N.J., 1959, pp. 479-484.

[2]Frank S. Freeman, The Fifth Mental Measurements Yearbook, p. 481.

[3]D.A. Pidgeon, The Fifth Mental Measurements Yearbook, p. 483.

[4]Ibid., pp. 483-84.

the ability to take tests and to understand classroom language, among other things. However, learned and innate mental capacities no doubt at least partially underly these abilities. The statistical measures and analyses were chosen to facilitate the accounting for differences in mental capacity of the groups which might affect reading and the other criterion skills measured by the posttests. In order to avoid any possible confounding of the results, an analysis of variance and an analysis of covariance using IQ as the covariable are reported.

Iowa Tests of Basic Skills[1]

This is a group test measuring skills in vocabulary, reading, language, work-study and arithmetic. The scores used in the Year Four (1967-68) analysis are from Form 4 for Grade Three and from Form 3 for Grade Four.

The battery actually consists of one long test for each area or sub-area, but each grade takes only a part of each test. The parts assigned to each grade level partially overlap each other.

Vocabulary:[2] Each item contains the stimulus word in the context of a phrase or sentence. The pupils must

---

[1]E.F. Lindquist and A.N. Hieronymus, Houghton Mifflin Co., Boston, 1964.

[2]The following descriptions of each test are adapted from Manual for Administrators, pp. 27-37.

mark the correct one of four definitions. Each definition is a word or a phrase.

Reading Comprehension: The pupils read several selections that vary in length up to half a page. Then the pupils answer questions based on the selections by marking one of four responses.

> The passages were chosen in an attempt to represent as completely as possible all of the types of material encountered by the pupil in his everyday reading. They were adapted from a wide variety of sources: newspapers, magazines, encyclopedias, government publications, textbooks and original literary works.[1]

Some of the skills measured in the Reading Comprehension test include:

(1) Recognizing and understanding stated or implied factual details and relationships.

(2) Discerning the purpose or main idea of a paragraph or selection.

(3) Organizing ideas.

(4) Evaluating what is read.

Language Skills: These tests consist of four parts: Spelling, Capitalization, Punctuation and Usage. On the Spelling test, each item has four words listed. One or none of the words is misspelled. The pupil marks either the incorrectly spelled word or a "No mistakes" response.

---

[1] Manual for Administrators, p. 27.

"The tests were constructed to sample systematically a wide variety of error types" such as final e or e before suffix, interchanged letters, and t and ed substitutions.[1]

In both the Capitalization and Punctuation tests each item has one or two sentences printed in three lines of about equal length. The pupil marks the line with an error or a fourth response labeled "No mistakes."

Each item in the Usage test contains three sentences. One or none of the sentences has a usage error in it. The pupil marks the sentence with the error or a fourth response labeled "No mistakes." The errors are such things as use of pronouns and verbs and avoidance of double negatives.

Work-Study Skills: These tests cover: (1) knowledge and use of map materials; (2) knowledge and use of graphic and tabular materials; and (3) knowledge and use of reference materials. In the Map Reading test each map is followed by several questions. The pupil must pick the correct answer from four responses. Various types of maps are included. The skills include such things as determining distance and describing location.

The test entitled Reading Graphs and Tables presents several types of graphs and tables. A series of questions requires the pupils to display such skills as

---

[1] Ibid., p. 32.

reading facts from the graphs or tables, making comparisons and interpreting. The pupil selects the correct response from four possibilities.

In Knowledge and Use of Reference Materials the pupil again chooses the correct answer from four possibilities. The topics include using word lists, using general reference materials, alphabetizing words, using an index, using a table of contents and using an encyclopedia.

Arithmetic Skills: These tests include items on (1) arithmetic concepts and (2) arithmetic problem solving. In both tests each item consists of a question. In Arithmetic Concepts the child has to pick one of four answers. In Arithmetic Problem Solving the child has to solve the problem stated in the item. The four responses he has to choose from include three possible answers and a response labeled "Not given." The skills in the arithmetic tests cover concepts involving currency, decimals, equations, fractions, geometry, measurement, numerals and number systems, percents, ratios and proportions, and whole numbers.

The number of items for each test at each grade and the time limits allowed for each test are given below:[1]

---

[1] Ibid., p. 5.

Test	Number of Items Grade Three	Number of Items Grade Four	Time in Minutes
Vocabulary	31	38	17
Reading Comprehension	60	68	55
Spelling	31	38	12
Capitalization	38	39	15
Punctuation	38	39	20
Usage	32	32	20
Map Reading	27	32	30
Reading Graphs and Tables	20	24	20
Knowledge and Use of Reference Materials	42	52	30
Arithmetic Concepts	30	36	30
Arithmetic Problem Solving	25	27	30

With regard to the norms used in this study, the publishers of the Iowa Tests of Basic Skills said that the sampling units for standardization purposes were selected primarily on the basis of community size and socio-economic characteristics.[1] No mention was made of the languages or dialects spoken by the children in the standardization sample.

---

[1] Ibid., p. 51.

The following conclusions from the reviews in The Fifth Mental Measurements Yearbook are especially relevant for this study:

(1) "The reviewer would strongly recommend the tests as a whole as the best of their kind."[1]

(2) Intercorrelation coefficients and examination of test items reveal that the skills most heavily tested are vocabulary and reading.

(3) "Experience suggests that, together with reading comprehension and vocabulary, the language tests should form a composite yielding a reasonable measure of language attainment."[2]

(4) Split-half reliability coefficients are quite satisfactory especially for the major test and composite scores.

Testing Procedures

The Lorge-Thorndike Intelligence Tests were given to the grade four children in the fall of 1966 and to the grade three children in the fall of 1967. The tests were administered and scored by the classroom teachers.

The Iowa Tests of Basic Skills were administered by the teachers to both grades in April, 1968. The tests

---

[1] G.A.V. Morgan, The Fifth Mental Measurements Yearbook, p. 36.

[2] Ibid., p. 35.

were electronically scored by the Houghton Mifflin Scoring Service in Iowa City, Iowa.

Data Analysis Procedures

The first run through the CDC 6600 Computer at The University of Texas at Austin Computation Center produced means and standard deviations for each experimental group in each grade on each of the tests: one set for The Lorge-Thorndike Intelligence Tests and fifteen for the Iowa Tests of Basic Skills. Jennings' routine LINEAR, a statistical computer program from the Edstat-J Library at the Computation Center, was used for analysis of variance and covariance on the second run through the computer.[1]

RESULTS OF DATA ANALYSES

Descriptive Statistics

The means and standard deviations for The Lorge-Thorndike Intelligence Tests and the Iowa Tests of Basic Skills are reported in Table 8 and Table 9.

The scores of the Iowa Tests of Basic Skills are the grade equivalents of raw scores. "The grade equivalent of a given raw score on any test indicates the grade level at which the typical pupil makes this raw

---

[1]E. Jennings, A Subroutine System for Data Processing in the Behavioral Sciences, The University of Texas, Austin, 1968.

Table 8

Means, Standard Deviations, F-Ratios and Probability Levels for Spring Administration of Iowa Tests of Basic Skills for Grade Four in Year Four (1967-68), San Antonio

	LCE (N = 29)		LCS (N = 60)		SAC (N = 21)		Variance		Covariance	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	F	P	F	P
I.Q.[1]	88.72	10.26	87.02	10.76	91.62	10.68	1.45	.240		
Vocabulary	3.4[2]	.8	3.3	.8	3.5	1.0	# .85	.430		
Reading	3.8	1.0	3.5	1.0	3.8	1.0	.78	.462	.35	.705
Spelling	4.6	1.1	3.7	1.0	4.0	1.3	#5.83	.004**		
Capitalization	4.0	1.0	3.7	1.2	3.8	1.3	.51	.604	.49	.614
Punctuation	4.5	1.5	3.8	1.3	4.0	1.4	2.13	.124	2.07	.131
Usage	4.3	1.3	3.8	1.3	3.8	1.5	2.83	.064	3.22	.044*
Total Language	4.4	1.1	3.8	1.0	3.9	1.2	3.05	.052	3.20	.045*
Maps	4.2	1.2	3.7	1.0	3.9	.9	1.99	.142	1.79	.172
Graphs	4.2	1.1	3.7	1.0	4.0	1.2	2.02	.137	1.51	.225
References	4.1	1.0	3.7	.7	4.0	1.0	2.31	.105	1.78	.175
Total Work-Study	4.2	.9	3.7	.7	4.0	.9	3.13	.048*	2.67	.074
Arith. Concepts	4.1	1.2	3.6	.7	3.9	1.1	3.17	.046*	2.95	.057
Arith. Problems	4.2	1.3	3.6	.9	3.8	1.1	2.55	.083	2.41	.095
Total Arithmetic	4.1	1.2	3.6	.7	3.8	1.0	3.18	.046*	3.07	.051
Composite	4.0	.9	3.6	.7	3.8	1.0	#2.33	.102		

#assumption of parallelness of regression lines was not met; therefore, only analyses of variance is presented.

* $p < .05$

** $p < .01$

[1] I.Q. as measured by The Lorge-Thorndike Intelligence Tests, administered in fall, 1966.

[2] All scores for Iowa Tests of Basic Skills are grade equivalents, i.e., 3.4 = third grade, fourth month.

Table 9

Means, Standard Deviations, F-Ratios and Probability Levels for Spring Administration of Iowa Tests of Basic Skills for Grade Three in Year Four (1967-68), San Antonio

	LCE (N = 109)		NOA-LCE (N = 59)		LCS (N = 40)		SAC (N = 68)		Variance		Covariance	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	F	P	F	P
I.Q.[1]	88.88	9.96	82.58	10.78	82.30	10.98	91.96	14.10	10.23	.001**		
Vocabulary	2.72[2]	.7	2.8	.7	2.5	.6	3.2	1.0	# 9.91	.001**		
Reading	2.9	.7	2.9	.7	2.6	.6	3.3	1.1	6.70	.001**	3.62	.014*
Spelling	3.1	1.2	3.3	1.2	2.8	1.0	3.8	1.2	7.82	.001**	5.96	.001**
Capitalization	3.0	.9	2.9	1.1	2.6	.8	3.8	1.3	#12.67	.001**		
Punctuation	3.3	1.0	3.2	1.2	2.9	1.0	4.0	1.5	# 8.43	.001**		
Usage	3.0	.9	2.7	1.0	2.6	1.0	3.8	1.3	14.88	.001**	7.69	.001**
Total Language	3.1	.8	3.0	1.0	2.7	.8	3.8	1.2	13.93	.001**	8.56	.001**
Maps	2.9	.7	2.9	.8	3.0	.8	3.5	1.1	# 9.41	.001**		
Graphs	3.0	.8	3.1	.9	2.8	.6	3.4	1.2	5.11	.002**	3.07	.028*
References	3.1	.7	3.1	.7	2.9	.6	3.4	1.0	# 4.12	.007**		
Total Work-Study	3.0	.6	3.0	.7	2.9	.5	3.5	1.0	# 8.16	.001**		
Arith. Concepts	3.2	.8	3.1	.8	3.0	.6	3.6	.9	5.78	.001**	2.32	.076
Arith. Problems	3.3	.8	3.3	.9	2.9	.7	3.6	1.0	4.41	.005**	2.24	.084
Total Arithmetic	3.2	.7	3.2	.8	2.9	.6	3.6	.9	6.53	.001**	3.03	.030*
Composite	3.0	.6	3.0	.6	2.7	.5	3.5	.9	#11.94	.001**		

assumption of parallelness of regression lines was not met; therefore only analyses of variance is presented

* $p < .05$

** $p < .01$

[1] I.Q. as measured by The Lorge-Thorndike Intelligence Tests, administered in fall, 1967.

[2] All scores for Iowa Tests of Basic Skills are grade equivalents, i.e., 2.7 = second grade, seventh month.

score."[1] For scores between 3.1 and 8.1, the first digit represents the grade and the second digit the end of the month in which the typical student makes that particular raw score. If the second digit is a zero, the score corresponds to the raw score of the typical pupil at the beginning of the grade. The publisher warns that grade equivalents below 3.1 "may not be interpreted literally as indicating grade and month" because they were not obtained from actual distribution of scores.[2]

> Rather, they were derived statistically on the assumption that pupils grow in the skills measured at the same rate in the first two grades . . . as in the third to eighth grades. This assumption is not strictly valid.[3]

Comparing Treatment Groups by Subtest

Fourth Grade: Although the SAC group mean score was highest on the IQ measure, the LCE group scored highest on 13 of 15 possibilities on the Iowa Tests of Basic Skills (ITBS). The LCS group IQ was nearly the same as that of the LCE group, but the LCS means on the ITBS were lowest on 14 of 15 possibilities. This generalization thus accurately describes the order of the mean

---

[1] Manual for Administrators, p. 13.

[2] Ibid.

[3] Ibid.

scores of the three treatment groups on the ITBS: LCE > SAC > LCS. Results of analysis of variance used to determine which of the differences are statistically significant are described in the next section.

Third Grade: The SAC group attained the highest mean scores on the IQ measure and on every test of the Iowa Tests of Basic Skills. The LCS group received the lowest mean score on 14 of 15 possibilities on the ITBS. The LCE and NOA-LCE means were in between, with neither consistently superior to the other. This generalization describes the order of the mean scores of the four treatment groups on the ITBS: SAC > LCE = NOA-LCE > LCS. Results of analysis of variance used to determine which of the differences are statistically significant are described in the next section.

Comparing Subtest Scores of Treatment Groups

To compare the relative standing of any group on the different tests of the Iowa Tests of Basic Skills one cannot use the grade equivalent scores. The amount of variation in scores is not the same on every test. One must use percentile rankings. The procedure utilized was to change the grade equivalents to percentiles by using tables giving end-of-year norms for school averages.[1]

---

[1]Manual for Administrators, pp. 73-74.

The children in the treatment groups represented several schools but for the purpose of comparing the scores in the several areas, the assumption was made that the children in each group attended the same school.

Beginning-of-year norms were based on results obtained in a standardization program in which the Iowa Tests of Basic Skills were administered in approximately 350 school buildings in grades three through six. Mid-year and end-of-year norms were then established by interpolation.[1]

Table 10 and Table 11 provide the percentile rankings for Grade Four and Grade Three. If a group had a percentile ranking of 7, for instance, this means that that group, considered as a school, scored higher than seven percent of all other schools, at that grade level, and for that test.

Fourth Grade: For the LCE group, the Language total score and subtest scores were the strongest scores (15 to 39). The Work-Study Skills were second strongest. Arithmetic was next while Reading (10) and Vocabulary (5) were the lowest.

For the LCS group the percentiles were all almost the same, ranging between four and fifteen. The Language Skills percentiles were the highest.

---

[1] Ibid., p. 17.

Table 10

Percentile Ranks of Grade-Equivalent Averages on Iowa Tests of Basic Skills for Grade Four in Year Four (1967-68), San Antonio[1]

Test	LCE[2]	LCS[3]	SAC[4]
Vocabulary	5	4	7
Reading	10	5	10
Spelling	38	9	14
Capitalization	15	9	11
Punctuation	33	10	15
Usage	39	15	15
Total Language	28	11	13
Maps	22	7	12
Graphs	22	7	15
References	17	7	13
Total Work-Study	20	7	13
Arithmetic Concepts	15	4	10
Arithmetic Problems	21	5	8
Total Arithmetic	16	5	8
Composite	14	6	10

[1]Taken from End-Of-Year (April 15) Norms for School Averages, Grade 4, Manual for Administrators, p. 74.

[2]LCE = Language-Cognition English

[3]LCS = Language-Cognition Spanish

[4]SAC = San Antonio Curriculum

Table 11

Percentile Ranks of Grade-Equivalent Averages on Iowa Tests of Basic Skills for Grade Three in Year Four (1967-68), San Antonio[1]

Test	LCE[2]	NOA-LCE[3]	LCS[4]	SAC[5]
Vocabulary	8	10	5	21
Reading	8	8	3	20
Spelling	14	19	8	44
Capitalization	12	10	4	50
Punctuation	19	15	7	62
Usage	15	7	5	50
Total Language	13	11	5	48
Maps	6	6	9	32
Graphs	9	13	5	27
References	11	11	6	24
Total Work-Study	8	8	6	30
Arithmetic Concepts	14	10	8	37
Arithmetic Problems	22	22	7	44
Total Arithmetic	15	15	6	39
Composite	10	10	5	30

[1]Taken from End-Of-Year (April 15) Norms for School Averages, Grade 3, Manual for Administrators, p. 73.

[2]LCE = Language-Cognition English

[3]NOA-LCE = Non-Oral-Aural until Year Four and then Language-Cognition English

[4]LCS = Language-Cognition Spanish

[5]SAC = San Antonio Curriculum

The pattern for the SAC group was similar to that of the LCE group. Language Skills had the highest percentiles and Work-Study Skills were second highest. Again Vocabulary and Reading Comprehension were lowest.

Considering all the groups together, the Language Skills received the highest percentiles and Reading Comprehension and Vocabulary the lowest. As a matter of fact, for all three groups no percentile rank for any subtest or total score was lower than Vocabulary.

Third Grade: The overall pattern for third grade was similar to the fourth grade in that Vocabulary and Reading Comprehension were again the weakest relatively. The Arithmetic Skills total percentile ranking was highest for LCE, NOA-LCE and LCS and second highest for the control (SAC) group. The second highest rankings for LCE and NOA-LCE were in Language Skills. As with the fourth grade, the LCS percentile rankings were all very close together (3rd to 9th percentiles).

In both grades, the highest percentile for an experimental treatment group was 39, for the fourth grade LCE group in Usage. Most of the percentile rankings were much lower.

Inferential Statistics

Fourth Grade

The major null hypothesis for fourth grade stated that after four continuous years the LCE, LCS and Control (SAC) groups will not differ significantly on fifteen subtest and total scores from the Iowa Tests of Basic Skills. The statistical findings for analysis of variance and analysis of covariance are reported on Table 8.

IQ: Analysis of variance revealed no significant differences in non-verbal IQ as measured by The Lorge-Thorndike Intelligence Tests. The mean IQ's were: SAC--91.62; LCE--88.72 and LCS--87.02.

Vocabulary: The results of analysis of variance revealed that there was no significant difference in Vocabulary scores among the treatment groups. Analysis of covariance was not appropriate because the assumption that the regression lines be parallel was not met. Figure 1 provides a pictorial display depicting the non-parallel nature of the regression lines. It appears that at IQ's above 90, the regular San Antonio curriculum was more effective while below 90, the LCE and LCS treatments were more effective.

The experimental oral language programs apparently did not result in differential growth in vocabulary as measured by the Iowa Tests of Basic Skills. Since the tests

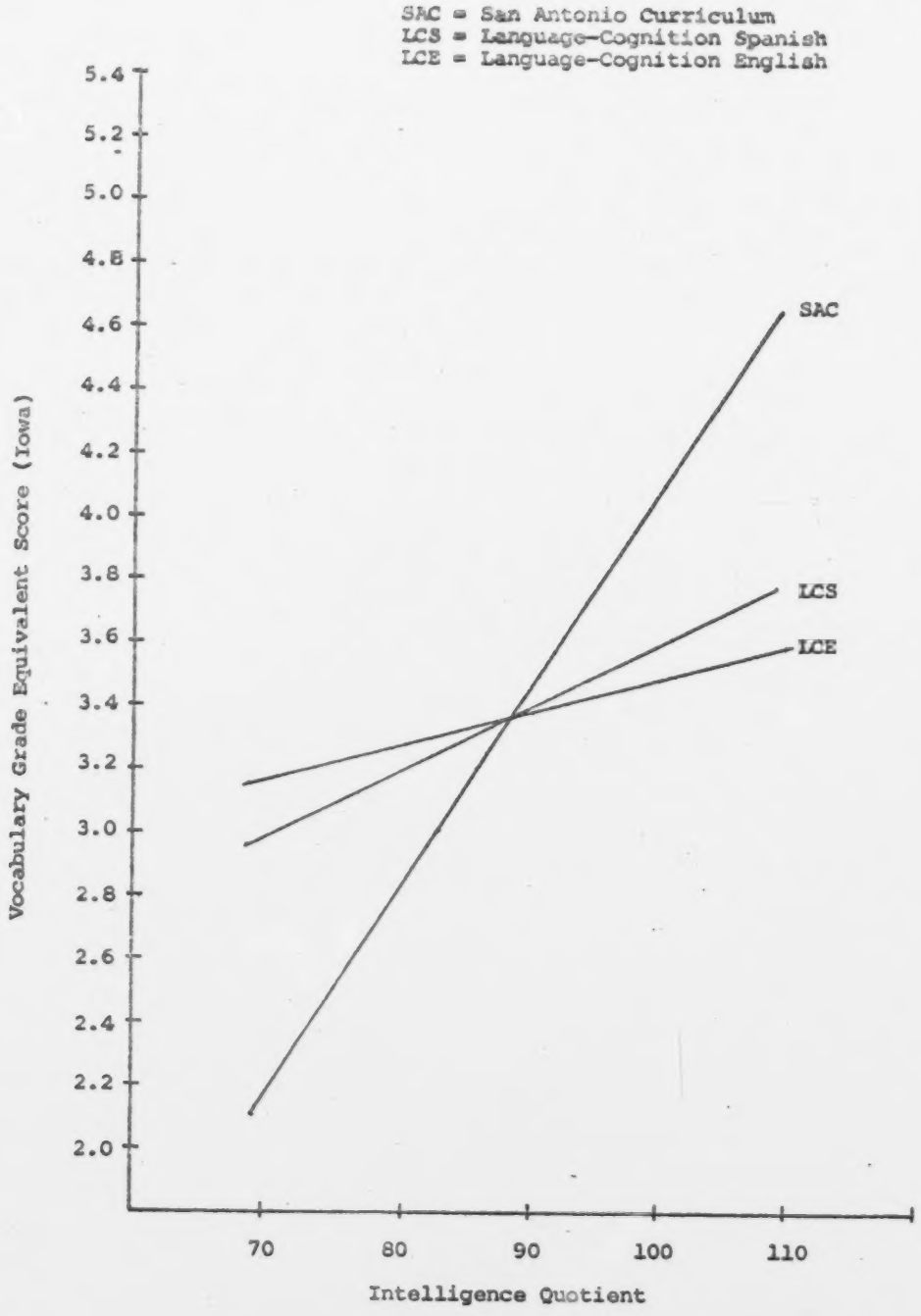

Figure 1. Least-squares Regression Lines for Each of the Three Groups (Grade Four, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Vocabulary Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

were administered in April, the typical child would have attained a score of 4.8 (fourth grade, eighth month). The mean scores for all three groups were more than a year below norm: SAC--3.5 (third grade, fifth month); LCE--3.4 and LCS--3.3.

Reading Comprehension: The results of analysis of variance showed that there was no significant difference in Reading Comprehension scores among the treatment groups. Analysis of covariance using IQ scores as the covariable also revealed no significant differences. The null hypothesis was accepted. The mean scores were: LCE--3.8; SAC--3.8 and LCS--3.5.

According to the results from this test, four years of intensive language training using the methods developed for the San Antonio Language Research Project did not produce differential growth in reading ability as measured by the Iowa Tests of Basic Skills.

Language Skills--Spelling: In Spelling, significant differences were found favoring the LCE group ($p = .004$). The mean score of the LCE group was 4.6. The LCS group mean was almost an entire year lower, 3.7. The control group (SAC) mean was 4.0. By using t-tests to compare each pair of means, it was found that the scores of the LCE and SAC groups were not significantly different, nor were those of the SAC and LCS groups. However, the LCE and LCS mean scores were significantly different at the .001 level.

Analysis of covariance was not appropriate because the regression lines were not parallel. Figure 2 provides a pictorial display showing the non-parallel nature of the regression lines. The LCE group performed much better than the LCS group at all IQ levels. Above IQ of 100, the SAC group scores were highest; between about 90 and 100, they were between the LCE and LCS scores; below 90, the SAC treatment was not as effective as the experimental language treatments.

On the basis of analysis of variance only, the null hypothesis was rejected. The results of this test indicate that the LCE treatment had a beneficial effect on development of spelling ability as measured by the Iowa Tests of Basic Skills.

Language Skills--Capitalization: According to both analysis of variance and analysis of covariance using IQ as the covariable, there was no significant difference in scores on the Capitalization subtest among the treatment groups. Therefore the null hypothesis was accepted. The mean scores were: LCE--4.0; SAC--3.8 and LCS--3.7.

If one of the groups receiving special language instruction had scored significantly higher or lower than the control group, it would have been difficult to supply an explanation. There was nothing in the experimental curriculum directed specifically at developing ability to capitalize correctly.

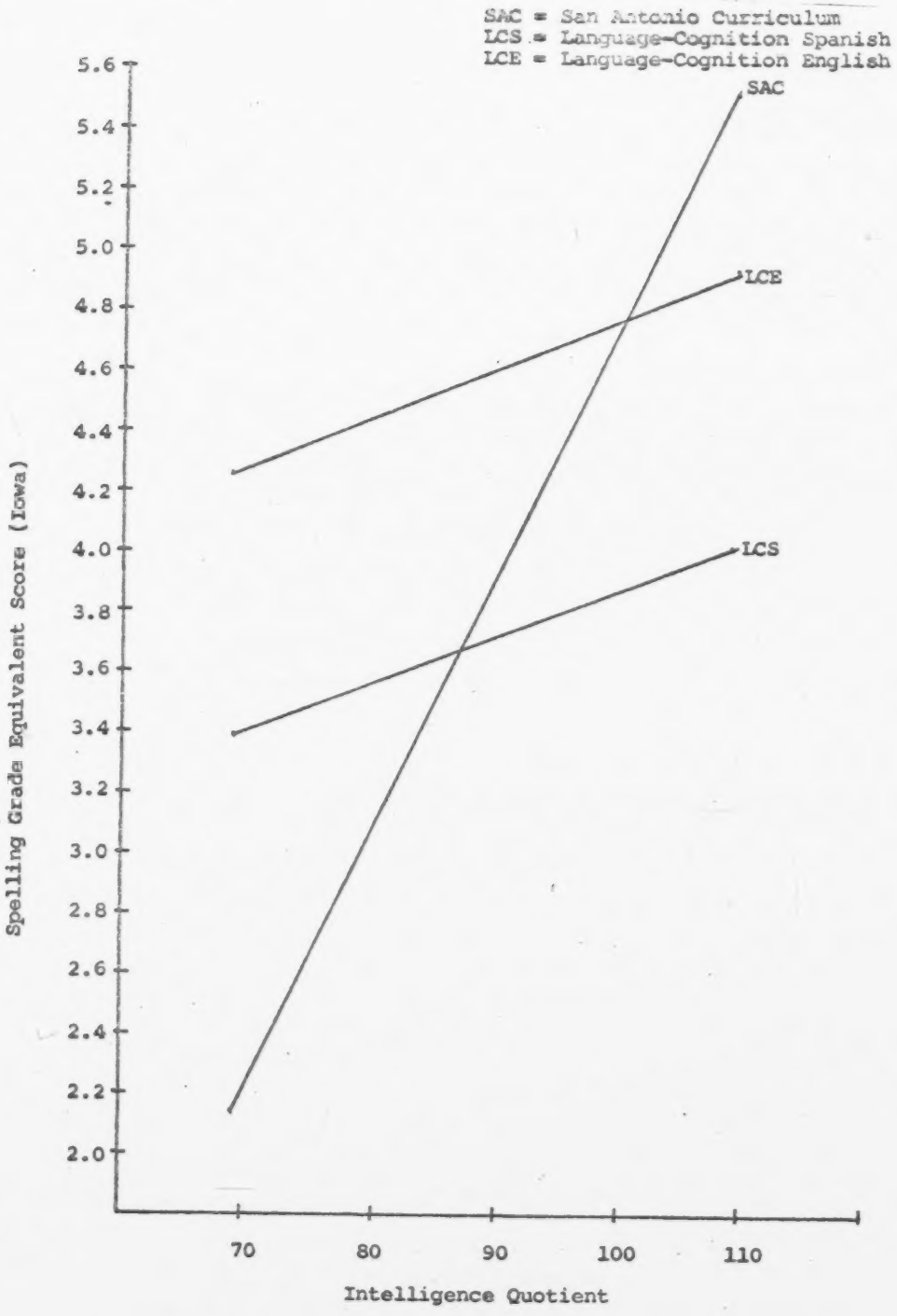

Figure 2. Least-squares Regression Lines for Each of the Three Groups (Grade Four, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Spelling Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

Language Skills--Punctuation: Analysis of variance and analysis of covariance using IQ as the covariable revealed no significant difference in scores on the Punctuation subtest among the treatment groups. Therefore the null hypothesis was accepted. The mean scores were: LCE--4.5; SAC--4.0 and LCS--3.8. Again, this result was expected. The language lessons had no content directed specifically at punctuation skills.

Language Skills--Usage: According to analysis of variance there was no significant difference in scores on the Usage subtest among the treatment groups. Thus, on the basis of analysis of variance, the null hypothesis was accepted. The mean scores were: LCE--4.5; SAC--3.8 and LCS--3.8.

According to analysis of covariance with IQ as the covariable, however, there was a significant difference ($p = .044$). Therefore, on the basis of analysis of covariance, the null hypothesis was rejected. Figure 3 graphically portrays the results of analysis of covariance. When scores were adjusted fo. IQ, the LCE treatment group scores remained highest at all IQ levels, but the LCS scores became second highest and the SAC scores were lowest. The order of these scores supports the hypothesis that the LCE treatment and perhaps the LCS treatment had beneficial effects on growth in language usage as measured by a written test, the Iowa Tests of Basic Skills.

SAC = San Antonio Curriculum
LCS = Language-Cognition Spanish
LCE = Language-Cognition English

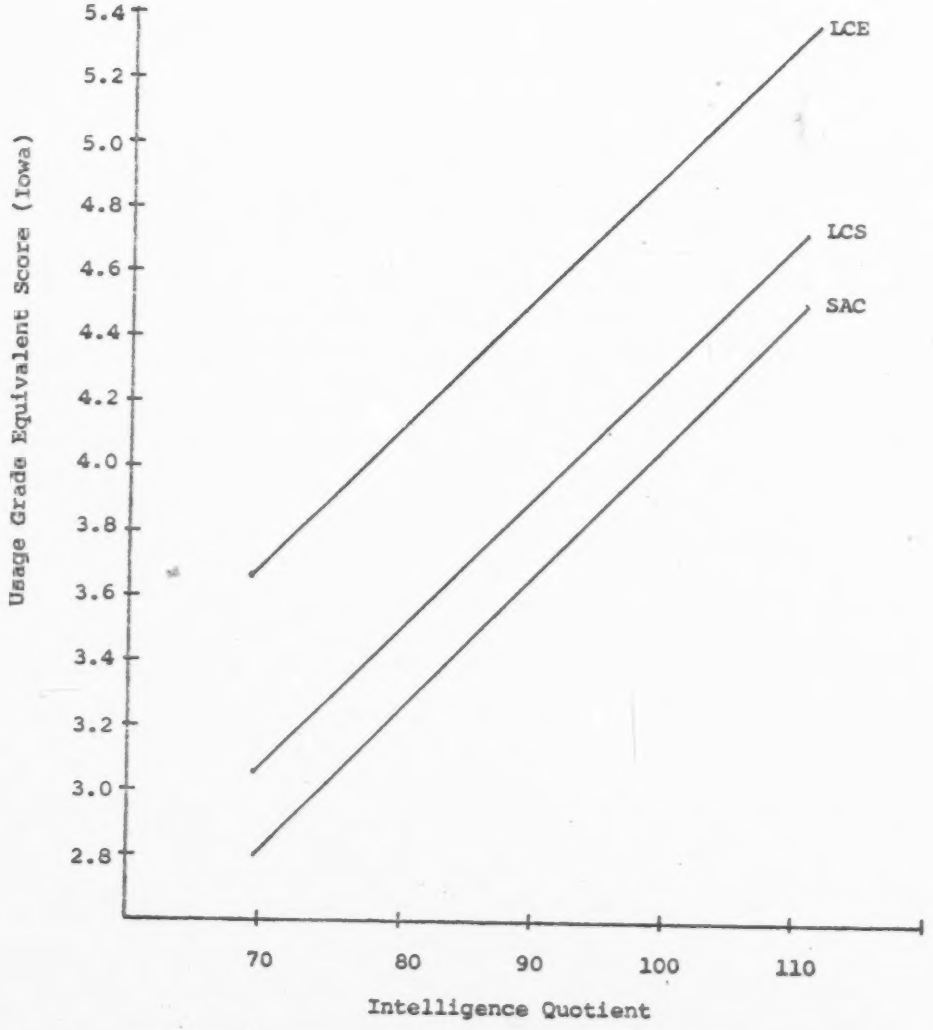

Figure 3. Least-squares Regression Lines for Each of the Three Groups (Grade Four, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Usage Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

Language Skills--Total: Analysis of variance revealed no significant differences in Language Skills--Total mean scores among the treatment groups, although the probability level did approach significance ($p = .052$). On the basis of analysis of variance, the null hypothesis was accepted. The mean scores were: LCE--4.4; SAC--3.9 and LCS--3.8.

When considering IQ as a covariable and using analysis of covariance, a significant difference was found among the mean scores of the Language Skills--Total subscale ($p = .045$). Thus, the null hypothesis was rejected. Figure 4 presents graphically the results of analysis of covariance. At all IQ levels, the LCE adjusted scores were almost six months higher than the LCS and SAC scores. The LCS and SAC scores were almost the same. This finding indicates that the LCE treatment had a beneficial effect on performance in the language area as measured by a written test, the Iowa Tests of Basic Skills.

Work-Study Skills--Map Reading: According to both analysis of variance and covariance, using IQ as the covariable, there was no significant difference in the scores on the Map Reading subtest among the three treatment groups. The null hypothesis was accepted. The mean scores were: LCE--4.2; SAC--3.9 and LCS--3.7. This result was not entirely expected since the fourth grade language lessons

SAC = San Antonio Curriculum
LCS = Language-Cognition Spanish
LCE = Language-Cognition English

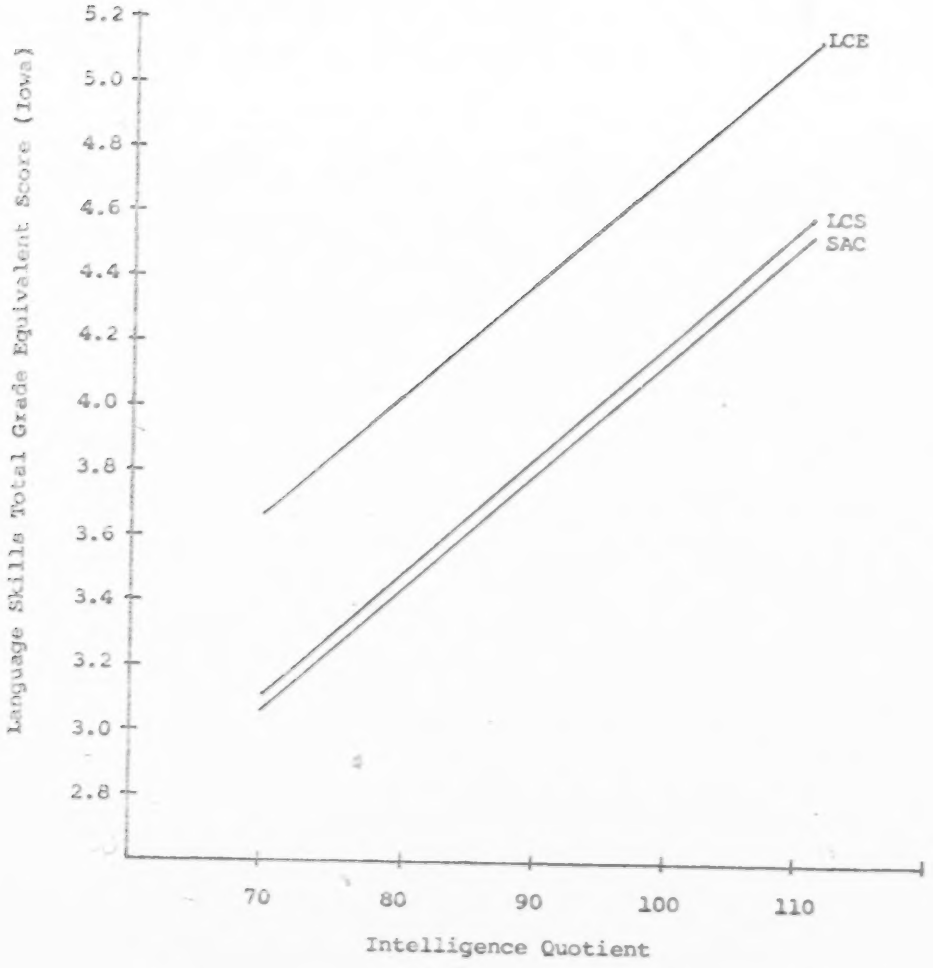

Figure 4. Least-squares Regression Lines for Each of the Three Groups (Grade Four, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Language Skills--Total Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

included activities involving construction and reading of maps. Unless the regular San Antonio curriculum included many of these kinds of activities the LCE and LCS should have shown greater strengths in this area. What was included in the regular San Antonio curriculum is not known.

Work-Study Skills--Reading Graphs and Tables: Both analysis of variance and of covariance revealed no significant differences in the mean scores of the three treatment groups on the subtest Reading Graphs and Tables. The null hypothesis was accepted. The mean scores were: LCE--4.2; SAC--4.0 and LCS--3.7.

Some of the AAAS activities in the LCE and LCS treatments involved the use of graphs and tables. Therefore it was expected that the LCE and LCS groups would have shown relative strengths on this subtest. Thus the results of analysis of variance and covariance were not those expected.

Work-Study Skills--Knowledge and Use of Reference Materials: Analysis of variance and analysis of covariance revealed no significant differences in the mean scores of the treatment groups. The null hypothesis was accepted. The mean scores were: LCE--4.1; SAC--4.0 and LCS--3.7.

Work-Study Skills--Total: According to analysis of variance there was a significant difference in Total scores of the Work-Study Skills tests ($p = .048$). On the basis of analysis of variance, the null hypothesis was rejected. The

mean scores were: LCE--4.2; SAC--4.0 and LCS--3.7. According to t-tests, neither the differences between the mean scores of the LCE and SAC groups nor between the SAC and LCS groups were significant. The difference between the LCE and LCS mean scores was significant at the .05 level.

According to analysis of covariance, however, there was no significant difference among the means. Thus on the basis of analysis of covariance the null hypothesis was accepted.

Since the AAAS activities involved use of both maps and graphs, it was expected that the LCE and LCS groups would have performed significantly better than the SAC group on the Maps and Graphs subtests and on the Work-Study Skills--Total. This differential growth was revealed for the LCE group only with analysis of variance as applied to the scores of Work-Study Skills--Total.

Arithmetic Skills--Arithmetic Concepts: The results of analysis of variance showed that there was a significant difference in the mean scores on the Arithmetic Concepts subtest ($p = .046$). Therefore, on the basis of analysis of variance, the null hypothesis was rejected. The mean scores were: LCE--4.1; SAC--3.9 and LCS--3.6. T-tests found significant differences only between the LCE and LCS mean scores ($p = .05$).

The findings favoring the LCE group according to analysis of variance may reflect the fact that some of the

oral language lessons included topics in arithmetic or arithmetic related concepts, e.g., shorter than, longer than, and thicker than. Results of analysis of covariance revealed no significant differences, but the probability level did approach significance ($p = .057$).

Arithmetic Skills--Arithmetic Problem Solving: According to both analysis of variance and analysis of covariance, there were no significant differences in the scores of the three treatment groups on the Arithmetic Problem Solving subtest. The significant difference favoring the LCE group found through analysis of variance in Arithmetic Concepts was not found in the scores on the Arithmetic Problem Solving subtest. The null hypothesis was accepted. The mean scores were: LCE--4.2; SAC--3.8 and LCS--3.6.

Arithmetic Skills--Total: On Arithmetic Skills--Total, significant differences were found favoring the LCE group using analysis of variance ($p = .046$). The null hypothesis was rejected. The mean scores were: LCE--4.1; SAC--3.8 and LCS--3.6. According to t-tests, the difference between the means of the LCE and LCS groups was significant at the .05 level, but the differences between the means of the LCE and SAC groups and between the SAC and LCS groups were not significant. When analysis of covariance was used, no significant differences were found in Arithmetic Skills--Total, and the null hypothesis was accepted.

Composite: Analysis of variance revealed no significant differences among the treatment groups on Composite mean scores. On the basis of analysis of variance, the null hypothesis was accepted. Mean scores were: LCE--4.0; SAC--3.8 and LCS--3.6. Analysis of covariance was inappropriate because the assumption that the regression lines be parallel was not met. Figure 5 depicts the non-parallel nature of the regression lines. The LCE scores were superior to the LCS scores at all IQ levels. At IQ's above about 95, the SAC scores were superior to both the special language treatments.

Thus, according to analysis of variance only, the LCE and LCS treatments did not produce differential growth in vocabulary, reading, language, work-study and arithmetic skills as summarized by a composite mean score on the Iowa Tests of Basic Skills.

In summary, according to analysis of variance, there were significant differences among the three treatment groups only for Spelling ($p = .004$), Work-Study Skills--Total ($p = .048$), Arithmetic Concepts ($p = .046$), and Arithmetic Skills--Total ($p = .046$). In all cases, the LCE group had the highest mean and the LCS group the lowest. The following generalization, based on analysis of variance, describes the differences, limited to those identified above as significant, between the groups: LCE > SAC > LCS, i.e., the LCE group was superior to the SAC group which was superior to the LCS group.

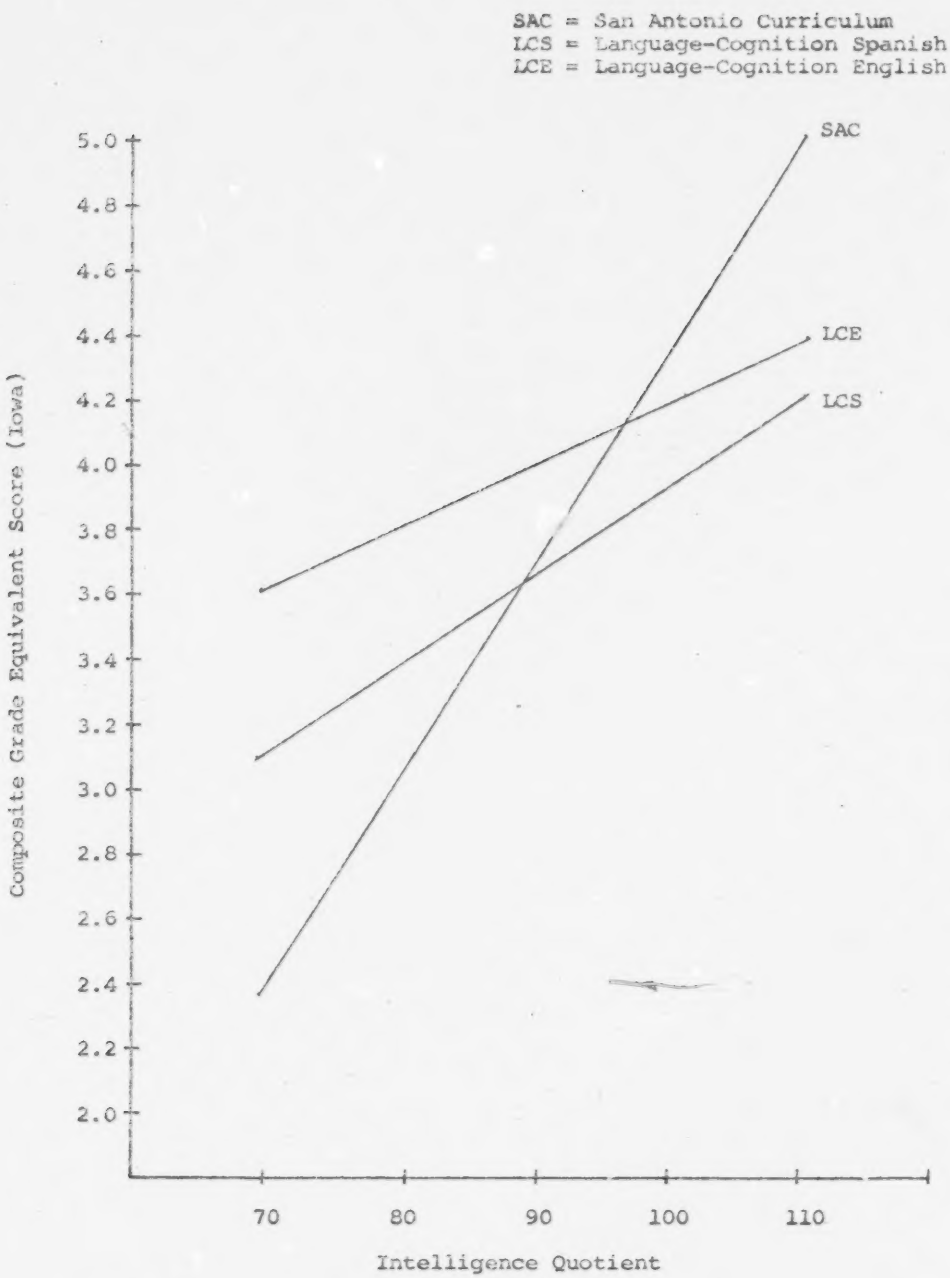

Figure 5. Least-squares Regression Lines for Each of the Three Groups (Grade Four, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Composite Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

According to analysis of covariance, there were significant differences among the groups only for Usage ($p = .044$) and Language Skills--Total ($p = .045$). The adjusted means were highest for the LCE group and lowest for the SAC group. The LCS adjusted means were much closer to the SAC mean scores. Thus, this limited generalization is appropriate: LCE > LCS ≅ SAC, i.e., the LCE group was superior to the LCS group which was only slightly superior to the SAC group for Language Usage and Language Skills--Total.

The figures depicting the regression lines for analysis of covariance for the Vocabulary, Spelling and Composite subscales seem to indicate that the LCE and LCS treatments were more effective for children with lower IQ's, while children with higher IQ's performed better in the regular San Antonio curriculum.

Third Grade

The major null hypothesis for third grade stated that after three continuous years the LCE, LCS and NOA-LCE and Control (SAC) groups will not differ significantly on fifteen subtest and total scores from the Iowa Tests of Basic Skills. The statistical findings for analysis of variance and analysis of covariance are reported in Table 9.

IQ: Analysis of variance revealed significant differences in non-verbal IQ as measured by The Lorge-Thorndike

Intelligence Tests ($p = .001$). The Control group (SAC) had the highest mean IQ score, 91.96, and the LCE group had the next highest, 88.88. The LCS and NOA-LCE groups had almost identical mean IQ scores, 82.30 and 82.58, respectively. According to t-tests, the differences between the SAC and LCE groups and between the NOA-LCE and LCS groups were not significant. All other differences (SAC vs. LCS; SAC vs. NOA-LCE; LCE vs. LCS; and LCE vs. NOA-LCE) were significant at the .001 level.

Vocabulary: The results of analysis of variance revealed significant differences in Vocabulary scores among the treatment groups ($p = .001$). Therefore the null hypothesis was rejected. The Control (SAC) group received the highest score, 3.2 (third grade, second month). The scores for the experimental groups were all a year or more below the scores for the typical pupil at this time of year: 3.8. The mean scores were: NOA-LCE--2.8; LCE--2.7 and LCS--2.5. T-tests revealed that the difference between the SAC group mean and the NOA-LCE mean was significant at the .01 level; that the differences between the SAC mean and the LCE and LCS means were each significant at the .001 level; that the difference between the NOA-LCE mean and LCS mean was significant at the .05 level; and that the differences between the LCS mean and both the NOA-LCE and LCE means were non-significant. The experimental language treatments clearly did not produce

differential growth in vocabulary, as measured by the Iowa Tests of Basic Skills, and not accounting for differences in IQ.

Analysis of covariance was not appropriate since the assumption that regression lines be parallel was not met. Figure 6 depicts the non-parallel nature of the regression lines. That this test could not be performed was unfortunate since one of the major goals of the language programs was to increase vocabulary. The figure showing the regression lines reveals that no generalization is possible at IQ's below about 100. Above 100, however, the order of the scores from highest to lowest is: SAC; NOA-LCE; LCE; and LCS. For the higher IQ's, at least, the experimental treatments did not have the hoped for beneficial effect on vocabulary development.

Reading Comprehension: Analysis of variance found significant differences in Reading Comprehension scores among the treatment groups ($p = .001$). Using analysis of variance, the null hypothesis was rejected. The mean scores were: SAC--3.3; LCE--2.9; NOA-LCE--2.9 and LCS--2.6. T-tests showed that the difference between the LCE and NOA-LCE means was not significant. Comparisons of all other differences between means showed them to be significant at the .05 level or higher.

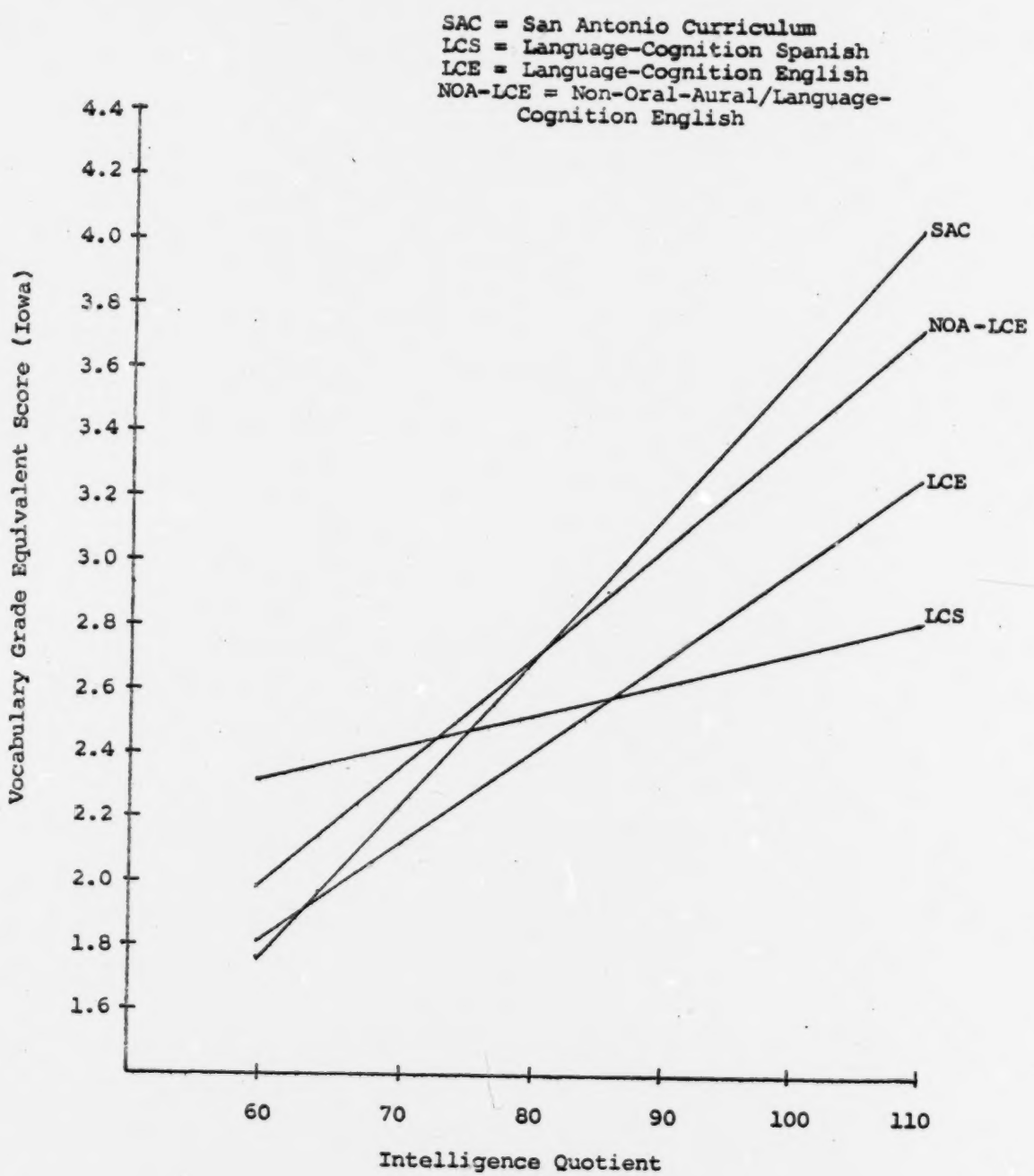

Figure 6. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Vocabulary Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

When IQ was controlled by using analysis of covariance, significant differences were found ($p = .014$). The null hypothesis was rejected. Figure 7 graphically portrays the results of analysis of covariance. When the scores were adjusted for IQ, the highest scores were still obtained by the SAC group; the scores of the NOA-LCE group drew ahead of those of the LCE group; and the lowest scores were still obtained by the LCS group. At this grade level, according to the Iowa Tests of Basic Skills, the hoped for effect on reading of the LCE and LCS treatments was not obtained.

Language Skills--Spelling: The results of analysis of variance revealed that there was a significant difference in the scores on the Spelling subtest of the four treatment groups ($p = .001$). On the basis of analysis of variance, the null hypothesis was rejected. The mean scores were: SAC--3.8; NOA-LCE--3.3; LCE--3.1 and LCS--2.8. According to t-tests, the differences between the SAC mean and the mean of every other group were significant at least at the .05 level. The differences between the NOA-LCE, LCE and LCS groups were not significant.

The findings of analysis of covariance also revealed significant differences ($p = .001$). Therefore, on the basis of analysis of covariance, the null hypothesis was rejected. Figure 8 portrays graphically the results of analysis of covariance. When scores were adjusted for IQ, the SAC group still received the highest scores and the NOA-LCE the second

SAC = San Antonio Curriculum
LCS = Language-Cognition Spanish
LCE = Language-Cognition English
NOA-LCE = Non-Oral-Aural/Language-Cognition English

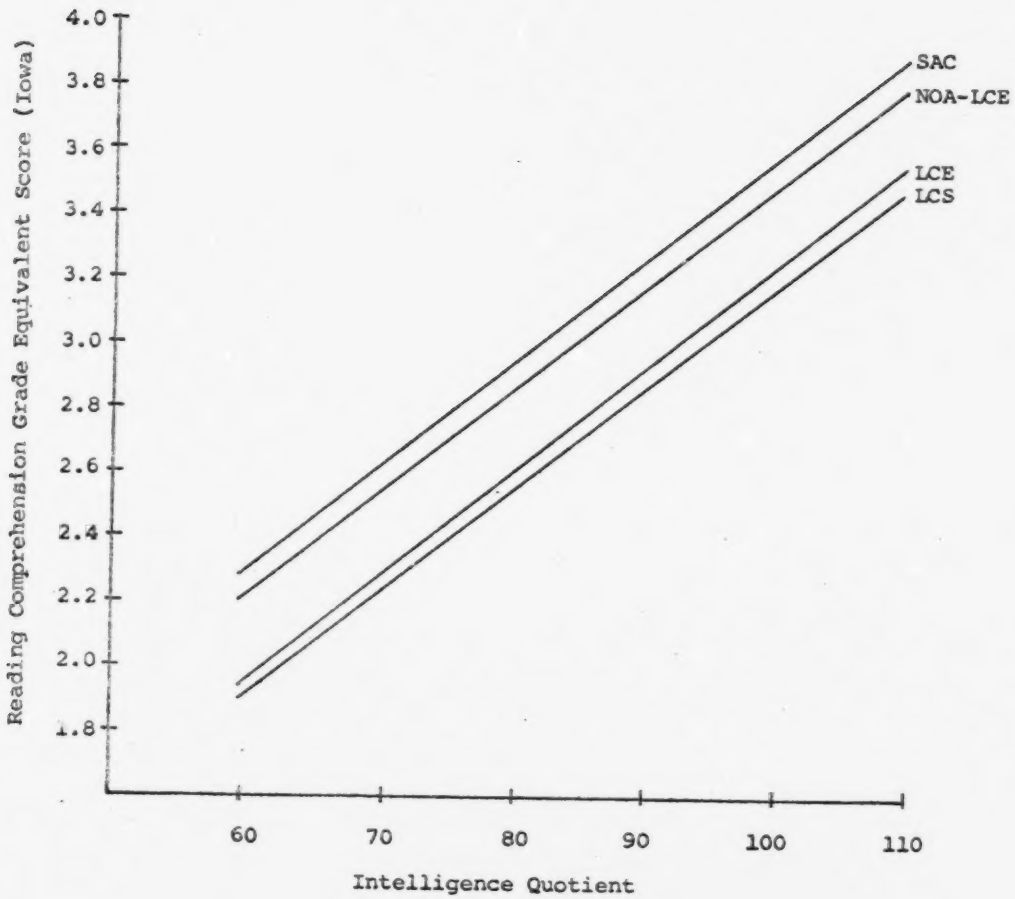

Figure 7. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Reading Comprehension Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

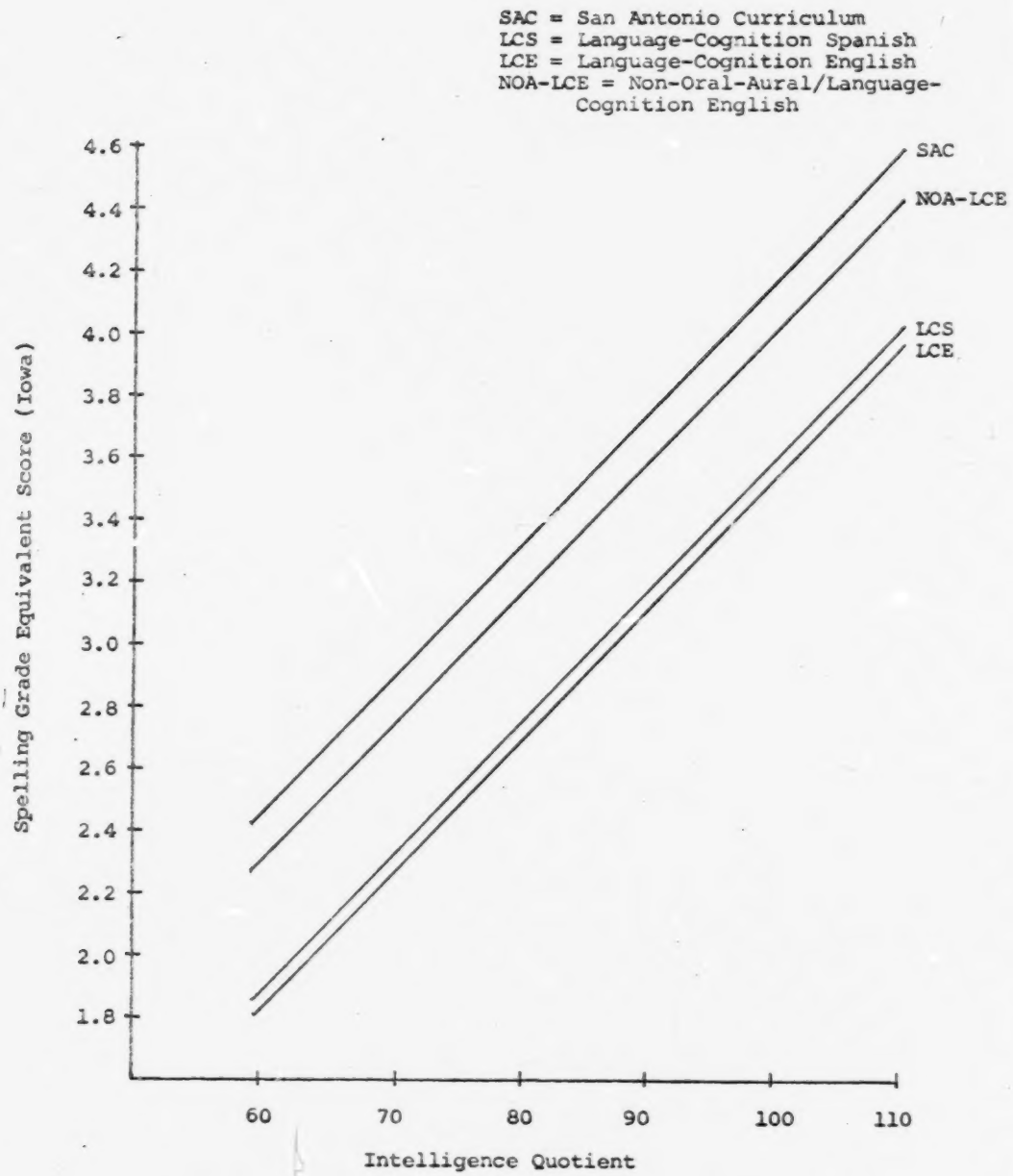

Figure 8. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Spelling Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

highest; the LCE and LCS received almost equal scores. The oral language programs in English and Spanish did not produce beneficial effects on spelling ability as measured by the Iowa Tests of Basic Skills.

Language Skills--Capitalization: According to analysis of variance, there were significant differences in the scores on the Capitalization subtest among the four treatment groups ($p = .001$). Again the SAC group scored the highest mean, 3.8, and the LCS group the lowest, 2.6. The difference between the highest and lowest mean was more than an entire year in grade equivalents. The LCE and NOA-LCE groups received almost the same score, 3.0 and 2.9, respectively. T-tests showed that this difference was not significant. The largest differences between means were between each experimental group and the control (SAC) group. According to t-tests, all these differences were significant at the .001 level. Also, the difference between the LCS mean and the LCE mean was significant at the .05 level, while the difference between the LCS and NOA-LCE means was not significant. There is no obvious feature of the experimental curriculum that provides an explanation for this pattern of scores.

On the basis of analysis of variance, the null hypothesis was rejected. Analysis of covariance was not appropriate because the regression lines were not parallel. Figure 9 depicts the non-parallel nature of the regression lines.

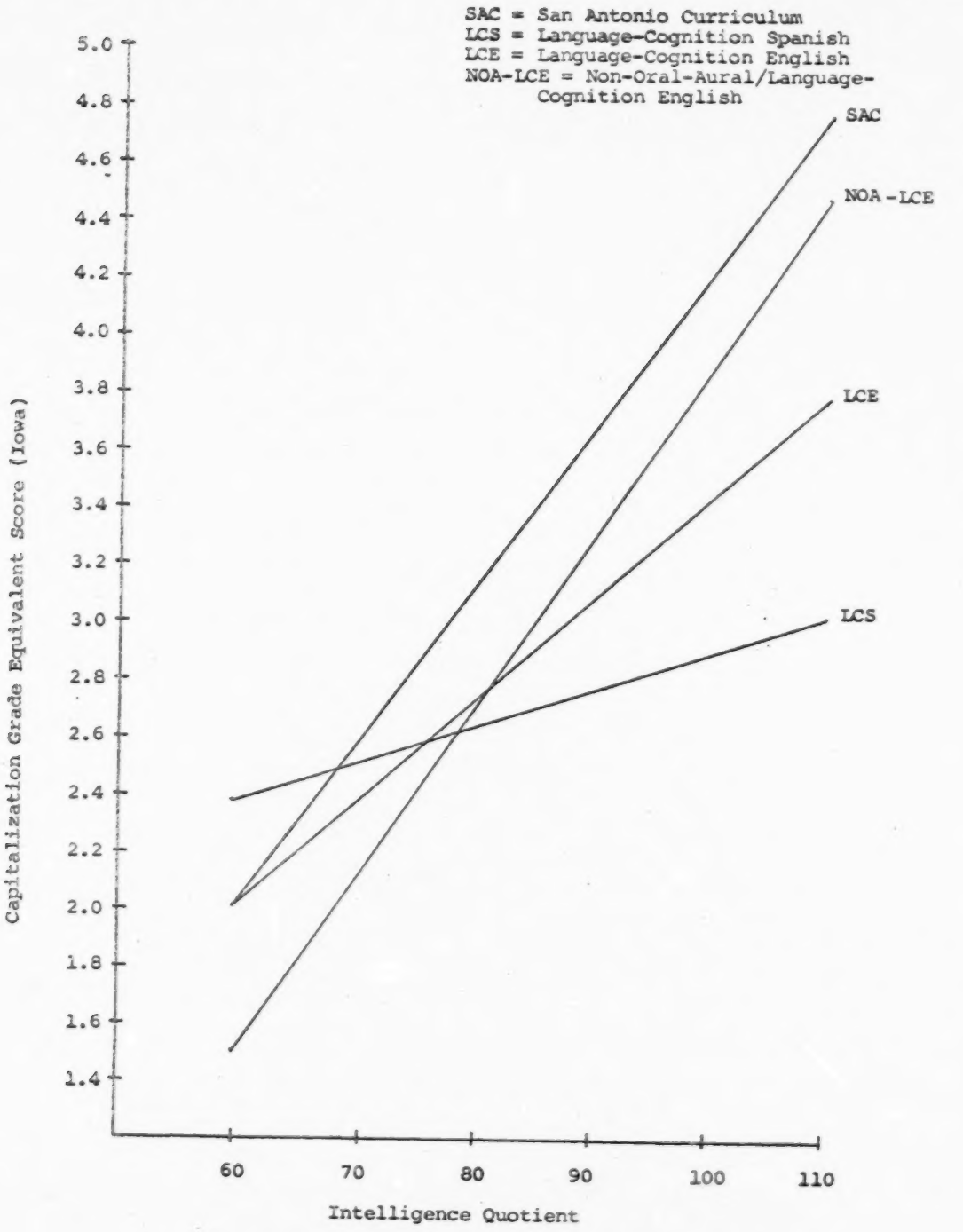

Figure 9. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Capitalization Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

Figure 9 shows that at IQ's above about 90, the typical ordering of scores occurs again. From highest to lowest they are: SAC, NOA-LCE, LCE and LCS. The differences are quite pronounced. Below 90, no generalization can be made about the scores of the four groups.

Language Skills--Punctuation: The results of analysis of variance revealed that there was a significant difference among the four treatment groups on the scores of the Punctuation subtest ($p = .001$). On this basis the null hypothesis was rejected. The mean scores were: SAC--4.0; LCE--3.3; NOA-LCE--3.2 and LCS--2.9. By applying t-tests to each pair of means, it was found that the differences between the SAC mean and the mean of each other group were significant at the .01 level or higher. The difference between the LCE and LCS means was significant at the .05 level. The differences between the NOA-LCE and both the LCE and LCS groups were non-significant. There is no obvious explanation for this pattern of scores on this particular test.

On the basis of analysis of variance, the null hypothesis was rejected. Controlling for IQ by using analysis of covariance was not appropriate since the assumption that the regression lines be parallel was not met.

Figure 10 shows the non-parallel nature of the regression lines. Above an IQ of 80, children achieved higher scores if they were in the SAC or NOA-LCE treatment groups.

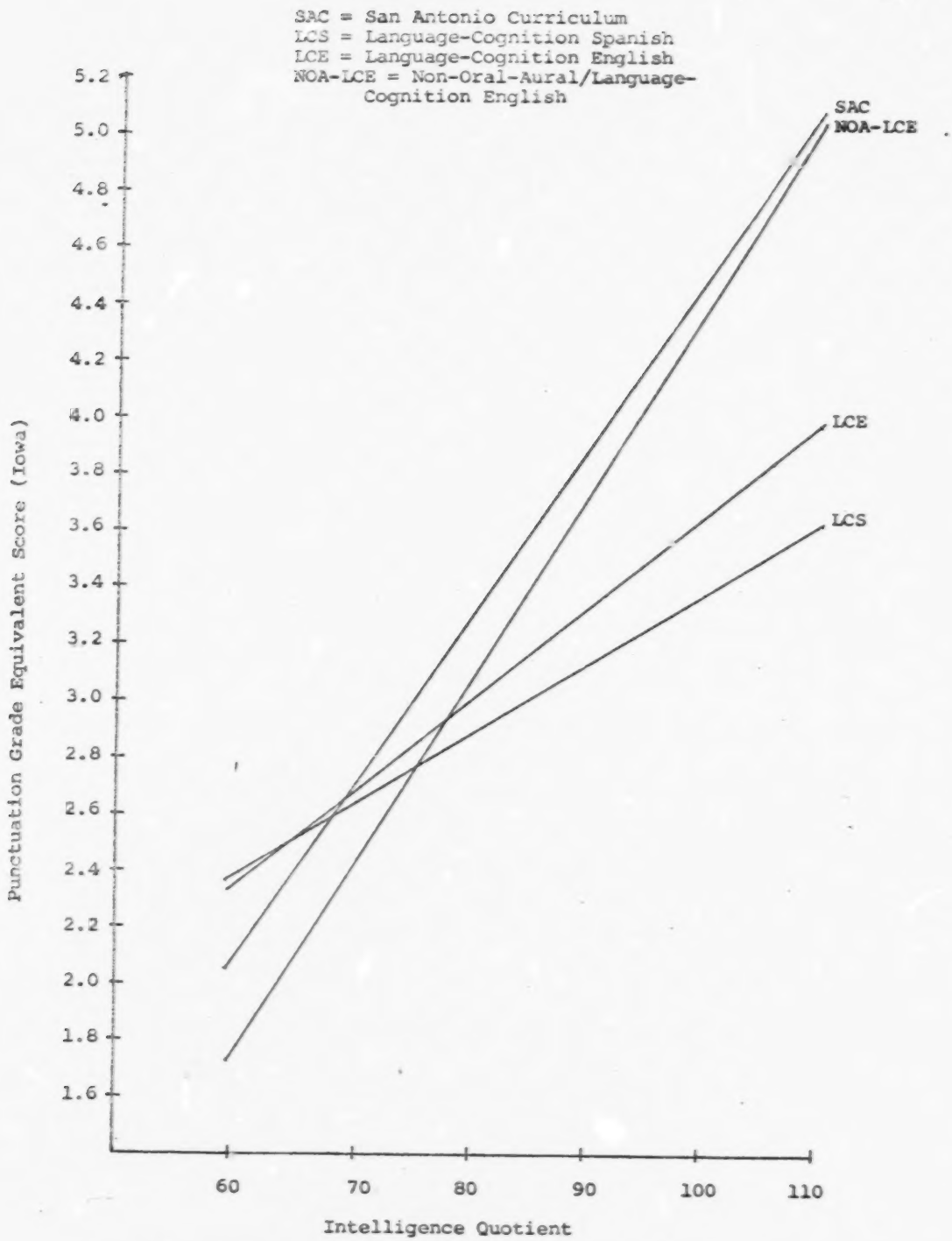

Figure 10. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Punctuation Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

Language Skills--Usage: The null hypothesis was rejected on the grounds that, both with analysis of variance and analysis of covariance, significant differences were found among the groups on the Usage subtest scores ($p = .001$). The mean scores were: SAC--3.8; LCE--3.0; NOA-LCE--2.7 and LCS--2.6. The differences between each of the experimental group means and the SAC group mean were significant at the .001 level. The differences between the LCE mean and the NOA-LCE and LCS means were significant at the .05 and .01 levels, respectively. The LCS and NOA-LCE mean scores were not significantly different.

Figure 11 depicts the results of analysis of covariance. When the scores were adjusted to account for IQ, the order of the groups remained the same except that the differences between the LCE and NOA-LCE scores disappeared. These findings provide evidence that the LCE and LCS treatments did not produce greater growth than the regular curriculum in language usage as measured by the Iowa Tests of Basic Skills.

Language Skills--Total: The results of both analysis of variance and of covariance showed that there were significant differences in the mean scores on Language Skills--Total among the four groups ($p = .001$). The mean scores were: SAC--3.8; LCE--3.1; NOA-LCE--3.0 and LCS--2.7. According to t-tests, the differences between the means of each experimental treatment group and the SAC mean were significant at

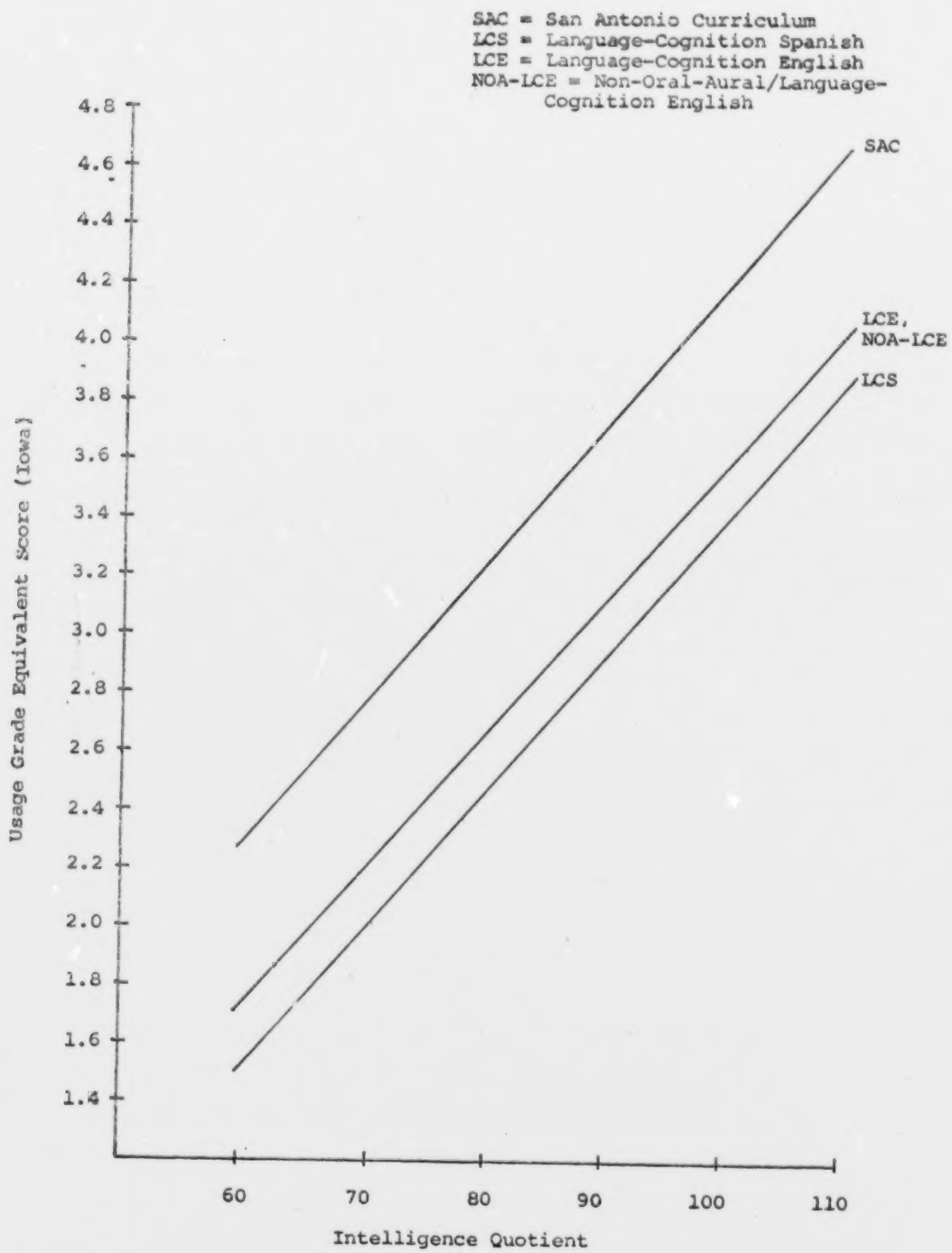

Figure 11. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Usage Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

the .001 level. The difference between the LCE and LCS means was significant at the .05 level. The NOA-LCE mean was not significantly different from either the LCE or the LCS mean.

Figure 12 presents graphically the results of analysis of covariance. When the scores were adjusted for IQ, the LCE group dropped from second place to third place, so again the children in the experimental language groups received the lowest scores.

The null hypothesis was rejected on the basis of both analysis of variance and of covariance. Language skills as measured by a written test, the Iowa Tests of Basic Skills, were not developed differentially by the LCE and LCS treatments.

Work-Study Skills--Map Reading: According to analysis of variance there were significant differences among the four groups on the Map Reading subtest ($p = .001$). The SAC group mean was highest: 3.5. The three experimental group means were about the same: LCS--3.0; LCE and NOA-LCE--2.9. T-tests showed that the differences between the SAC mean and the NOA-LCE, LCE and LCS means were significant at the .001, .001 and .05 levels, respectively. There were no significant differences between each combination of LCE, LCS and NOA-LCE group means. On the basis of analysis of variance, the null hypothesis was rejected.

Analysis of covariance was not appropriate because the regression lines were not parallel. Figure 13 displays the non-parallel nature of the regression lines. At very

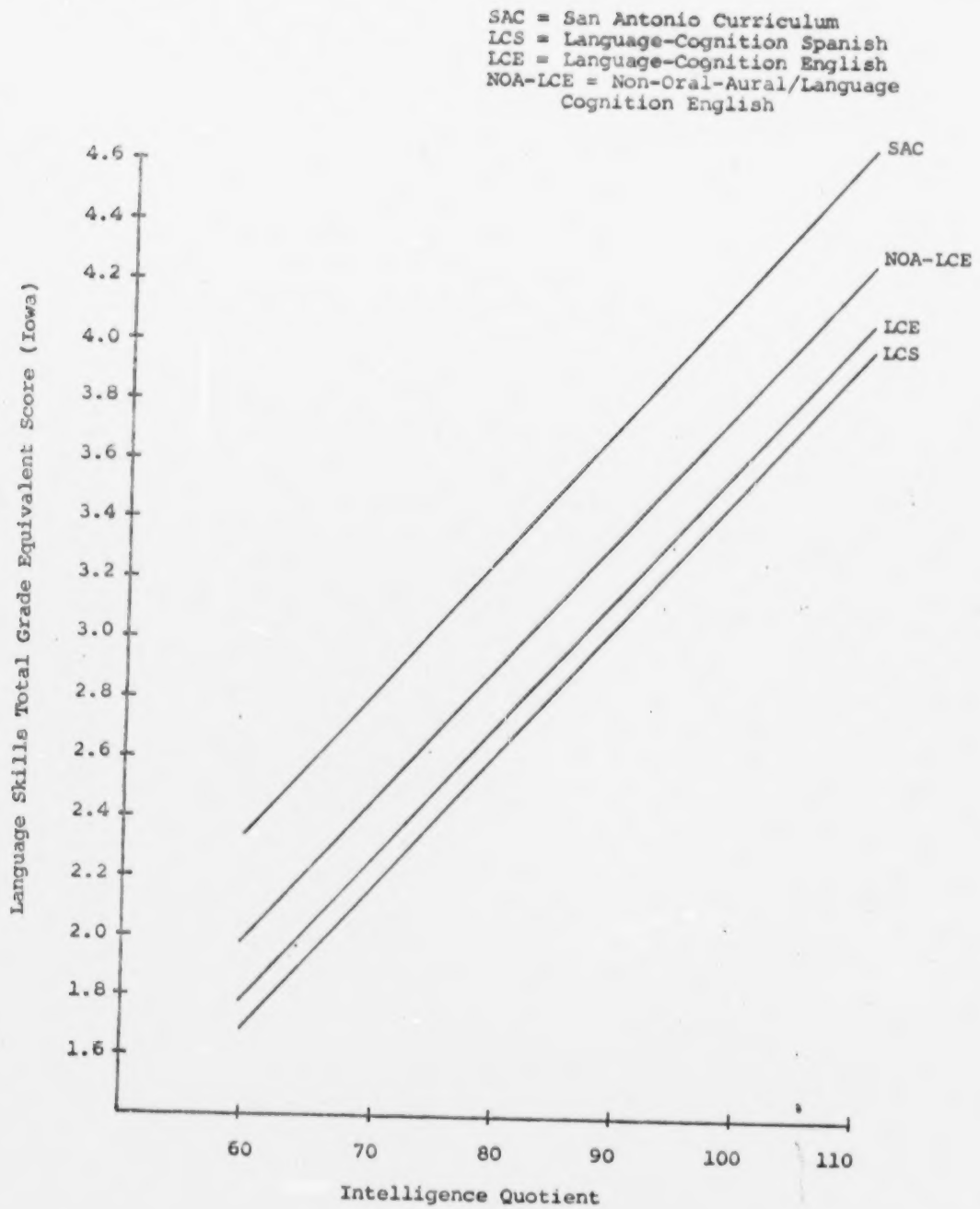

Figure 12. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Language Skills--Total Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

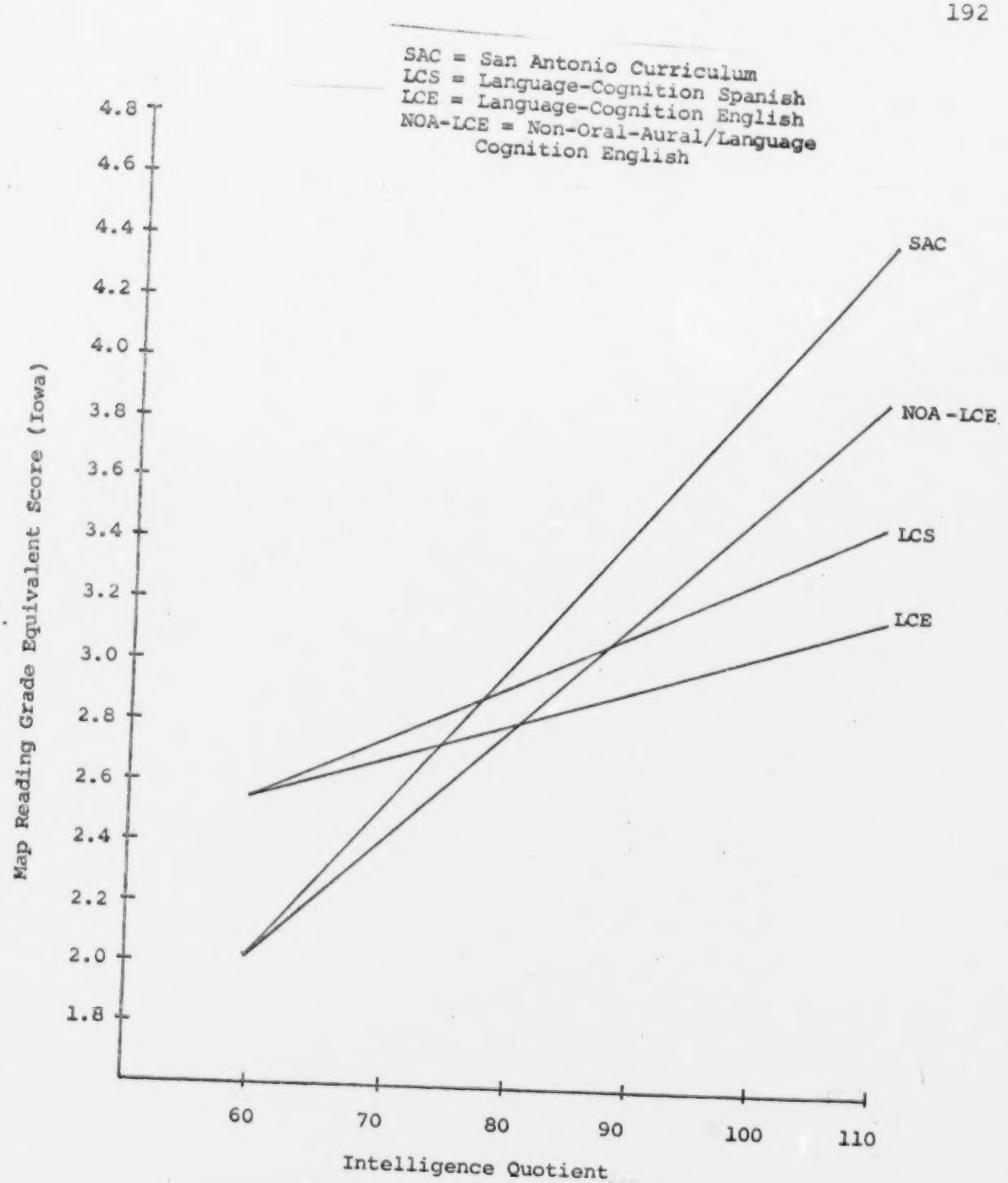

Figure 13. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Map Reading Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

low IQ levels, the experimental language treatments (LCE and LCS) were superior. Above an IQ of 90, the regular San Antonio curriculum and NOA-LCE treatment were superior.

Work-Study Skills--Reading Graphs and Tables: The results of analysis of variance showed that there were significant differences among the four groups on the Graphs subtest ($p = .002$). On the basis of analysis of variance, the null hypothesis was rejected. The mean scores were: SAC--3.4; NOA-LCE--3.1; LCE--3.0 and LCS--2.8. Again the differences between the SAC means and those of the NOA-LCE, LCE and LCS groups were significant according to t-tests. The probability levels were .05, .05 and .01, respectively. The differences between the NOA-LCE and the LCE groups, between the NOA-LCE and LCS groups, and between the LCE and LCS groups were not significant.

According to analysis of covariance, there were significant differences among the groups on the scores of the Graphs and Tables subtest. Figure 14 presents graphically the results of analysis of covariance. When the scores were adjusted for IQ, it was found that the SAC and NOA-LCE groups scored the same and the LCE and LCS groups received almost equal scores about three months lower than the other two groups. Apparently the experiences involving graphing in the science lessons were not effective in developing superior skill in using graphs in the LCE and LCS groups as measured by the Iowa Tests of Basic Skills.

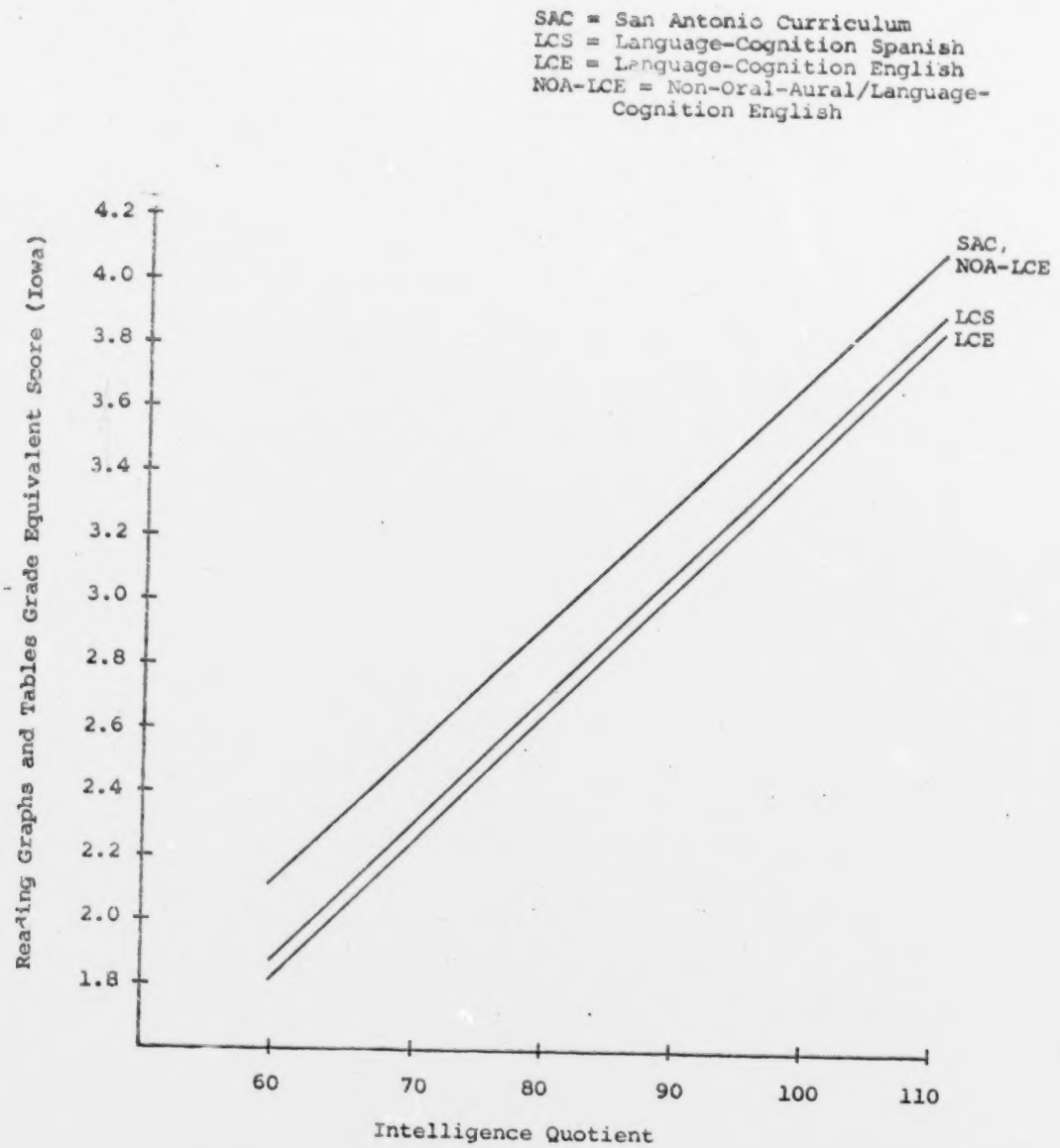

Figure 14. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Reading Graphs and Tables Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

Work-Study Skills--Knowledge and Use of Reference Materials: Analysis of variance found significant differences in the scores on the References subtest ($p = .007$). On the basis of analysis of variance, the null hypothesis was rejected. The mean scores were: SAC--3.4; LCE--3.1; NOA-LCE--3.1 and LCS--2.9. T-tests revealed that there were no significant differences between each pair of means for the experimental treatment groups. The difference between the SAC mean and NOA-LCE mean was significant at the .05 level, as was the difference between the SAC mean and the LCE mean; the difference between the SAC mean and the LCS mean was significant at the .01 level.

Analysis of covariance was not appropriate because the regression lines were not parallel. Figure 15 depicts the non-parallel nature of the regression lines. At IQ's above 100, the SAC and NOA-LCE treatments were superior to the LCE and LCS treatments. Below 100, the interaction was so great that it is impossible to generalize.

Work-Study Skills--Total: According to analysis of variance there were significant differences in the scores among the four groups. Therefore the null hypothesis was rejected. The mean scores were: SAC--3.5; LCE--3.0; NOA-LCE--3.0 and LCS--2.9. According to t-tests, the differences between the SAC group mean and the mean of the NOA-LCE, LCE and LCS groups were significant at the .01 level or higher.

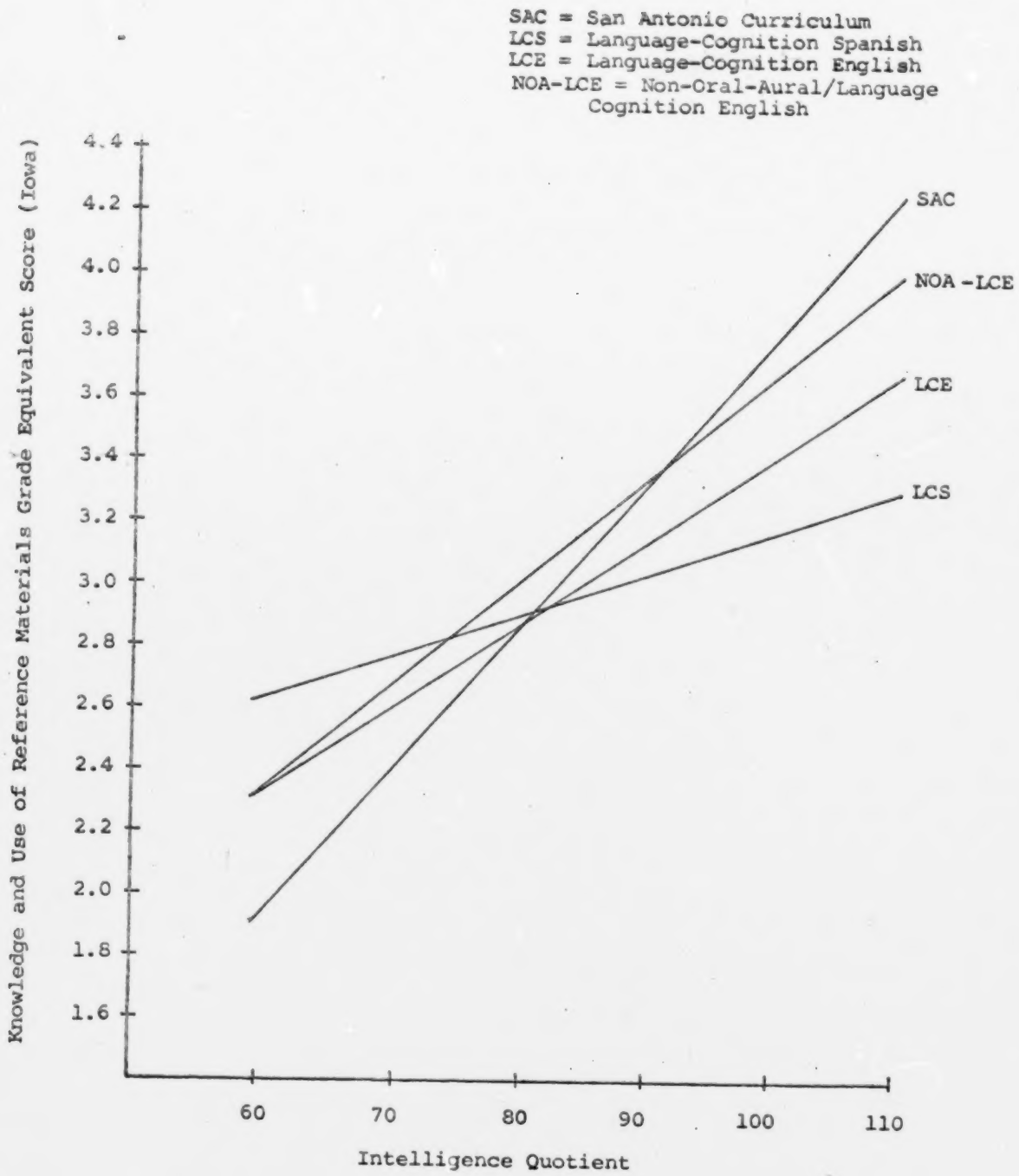

Figure 15. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Knowledge and Use of Reference Materials Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

The differences between the means of the LCE and NOA-LCE groups, between the LCE and LCS groups, and between the NOA-LCE and LCS groups were not significant.

Analysis of covariance was not appropriate because the regression lines were not parallel. Figure 16 depicts the non-parallel nature of the regression lines. As happened before at the higher IQ levels, the order of the groups from highest to lowest was: SAC, NOA-LCE, LCE and LCS. Below about an IQ of 90, the interaction prevents any generalization from being made.

Arithmetic Skills--Arithmetic Concepts: Analysis of variance found significant differences among the groups on the scores of the Arithmetic Concepts subtest ($p = .001$). The null hypothesis was rejected. The mean scores were: SAC--3.6; LCE--3.2; NOA-LCE--3.1 and LCS--3.0. Again the differences between the SAC group mean and the mean of each of the experimental groups were significant. The probability levels were .01 for the NOA-LCE and LCE groups and .001 for the LCS group. When the experimental group means were compared, no significant differences were found.

Analysis of covariance was performed but no significant differences were found ($p = .076$). Therefore the null hypothesis was accepted.

Arithmetic Skills--Arithmetic Problem Solving: There were significant differences among the groups on the scores

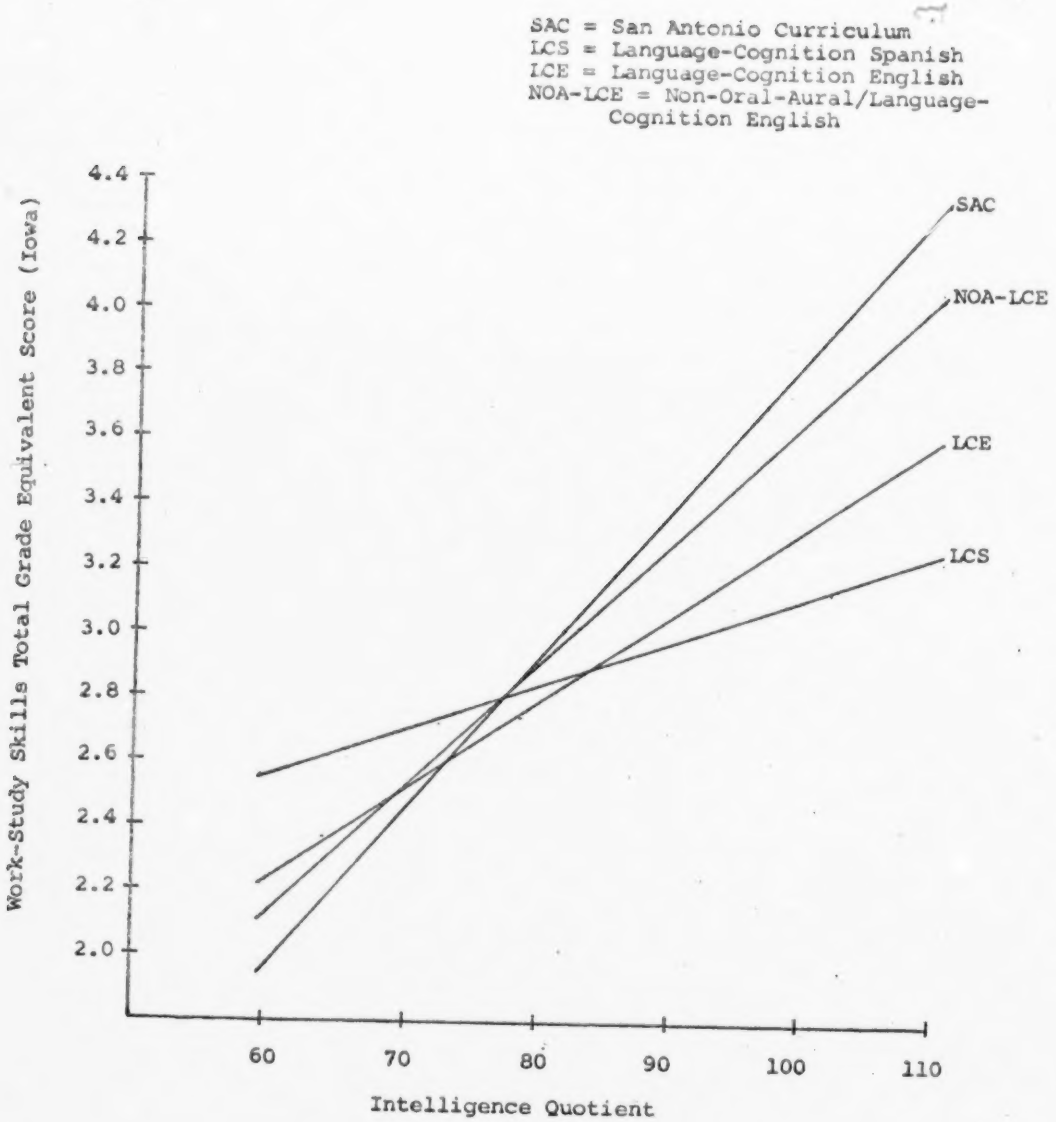

Figure 16. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Work-Study Skills--Total Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

on Arithmetic Problem Solving according to analysis of variance ($p = .005$). Therefore the null hypothesis was rejected. The highest score was obtained by the SAC group, 3.6; the lowest by the LCS, 2.9. According to t-tests, the difference between these two scores was significant at the .001 level. The LCE and NOA-LCE groups scored 3.3. These scores were not significantly different from the SAC mean, but each was significantly different from the LCS mean at the .05 level. Analysis of covariance found no significant differences ($p = .084$) so the null hypothesis could not be rejected on the basis of analysis of covariance.

The two Arithmetic Skills subtests are the only two tests for which analysis of covariance was appropriate and for which significant differences were not found. In all other cases the SAC group obtained the highest mean scores or were equal to the highest. This difference in the pattern of scores means that it is possible that the exercises involving mathematics within the AAAS materials were effective in producing relatively greater growth in arithmetic skills.

Arithmetic Skills--Total: According to analysis of variance there were significant differences in the mean scores of the four groups ($p = .001$). On the basis of analysis of variance, the null hypothesis was rejected. The mean scores were: SAC--3.6; LCE--3.2; NOA-LCE--3.2 and LCS--2.9. The SAC mean score was significantly different from the NOA-LCE, LCE and LCS means at the .05, .01 and .001 levels, respectively.

The mean of the NOA-LCE group was not significantly different from the LCE and LCS group means, but the LCE group mean was significantly different from the LCS mean at the .05 level.

According to analysis of covariance, there were significant differences when IQ was controlled ($p = .030$). Figure 17 depicts the results of analysis of covariance. When the scores were adjusted for IQ, the NOA-LCE scores were almost as high as those of the SAC group. The order of the other groups remained the same. The null hypothesis was rejected.

Composite: Analysis of variance applied to the Composite scores revealed significant differences among the groups' scores ($p = .001$). As with all the other tests, the SAC group obtained the highest mean: 3.5. Even this mean score is somewhat below that obtained by the typical child in the third grade at that time of year: 3.8. The other mean scores were: LCE--3.0; NOA-LCE--3.0 and LCS--2.7. As before, the differences between the SAC mean and the means of each other group were significant. In each case, the probability level was .001. The LCE and NOA-LCE mean scores were not significantly different from each other. The means of both the LCE and NOA-LCE groups were significantly different from the LCS mean at the .05 level.

Analysis of covariance was not appropriate because the regression lines were not parallel. Figure 18 portrays

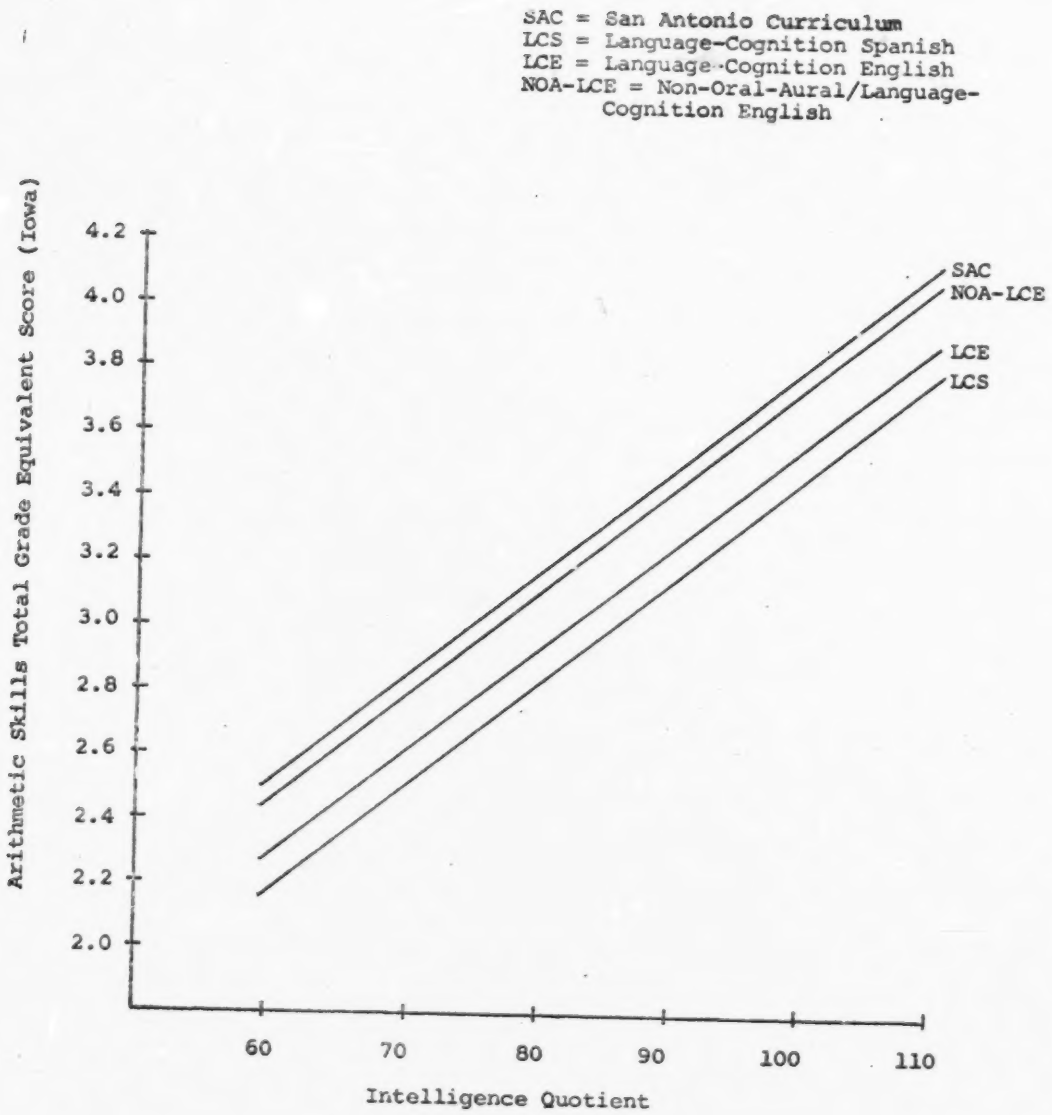

Figure 17. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Arithmetic Skills--Total Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

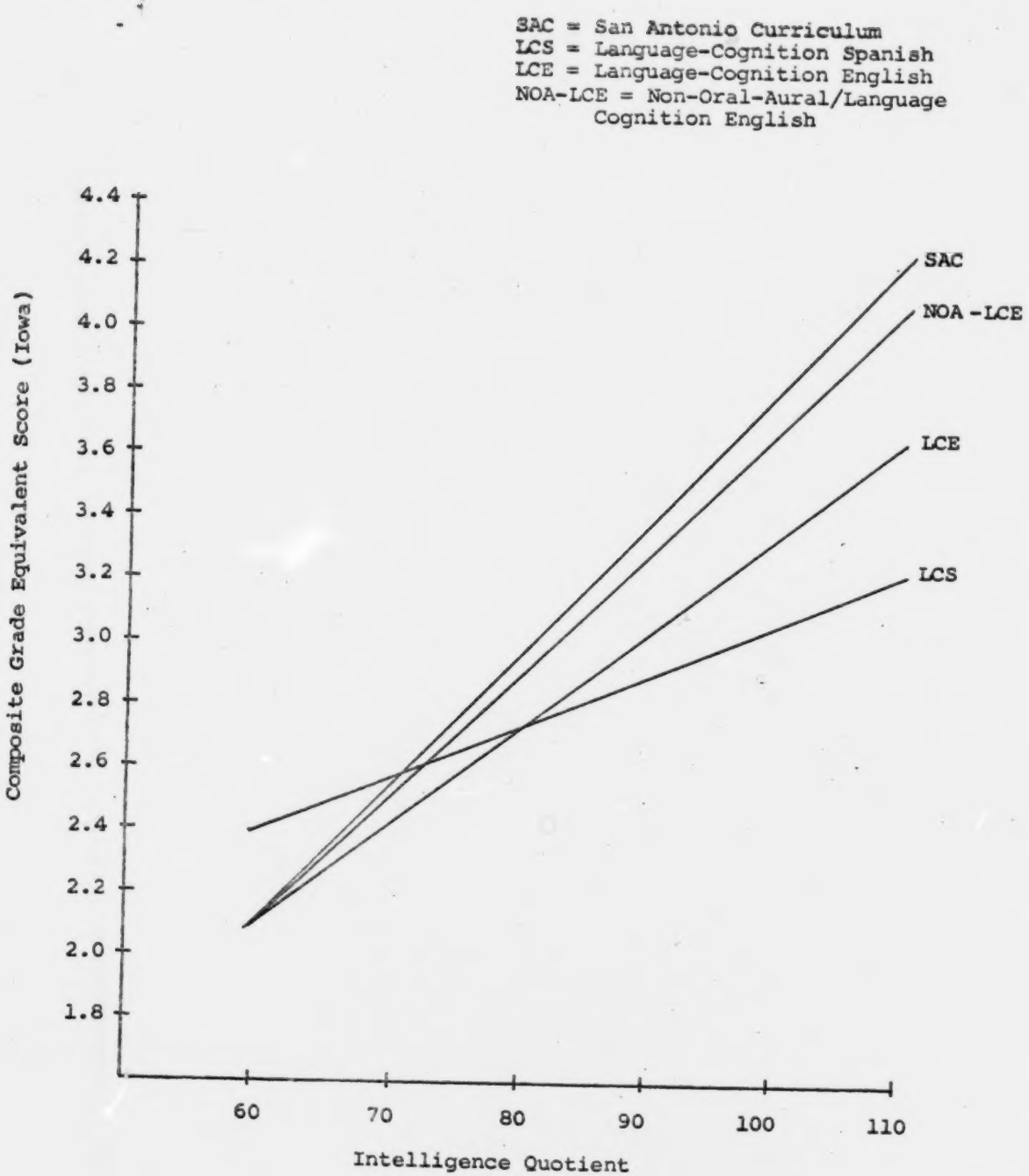

Figure 18. Least-squares Regression Lines for Each of the Four Groups (Grade Three, Year Four, 1967-68, San Antonio) Indicating the Relationship Between The Lorge-Thorndike Intelligence Tests Scores and the Iowa Tests of Basic Skills (Composite Subscale) Grade Equivalence Scores. These lines were extended approximately two standard deviations from the Grand Mean IQ.

the non-parallel nature of the regression lines. Above about an IQ of 85, the SAC and NOA-LCE treatments were clearly superior to the LCE treatment which was in turn superior to the LCS treatment.

In summary, according to analysis of variance, there were significant differences among the groups on IQ as measured by The Lorge-Thorndike Intelligence Tests and on every test of the Iowa Tests of Basic Skills. As can be seen on Table 9, the probability values are less than .01 in all cases. The control (SAC) group always obtained the highest mean scores. Except on the Reading Maps subtest, the LCS group always received the lowest mean scores. The mean scores of the LCE and NOA-LCE groups were between the mean scores of the SAC and LCS groups. Neither the LCE nor the NOA-LCE group was consistently superior to the other. Thus the following generalization describes quite accurately the differences between the four groups when IQ was not taken into consideration: SAC > LCE = NOA-LCE > LCS.

Given the significant differences in IQ scores, it was unfortunate that analysis of covariance was appropriate for only eight subtests of the Iowa Tests of Basic Skills. Significant differences were found in six of these cases: Reading Comprehension ($p = .014$), Spelling ($p = .001$), Usage ($p = .001$), Language Skills--Total ($p = .001$), Reading Graphs and Tables ($p = .028$) and Arithmetic Skills--Total ($p = .030$). The adjusted mean scores of the SAC group were always the

highest except in one case (Reading Graphs and Tables) when they were equalled by the scores of the NOA-LCE group. This latter group had the second highest adjusted mean scores in all other cases except on the Usage subtest on which the LCE mean score was higher. The adjusted mean scores of the LCE and LCS groups had the third and fourth highest scores in five of the six cases where there were significant differences according to analysis of covariance. The LCS group mean scores never rose above third highest. No significant differences were found for Arithmetic Concepts and Arithmetic Problem Solving subtests. The results according to analysis of covariance may be summarized generally, limited to the significant differences noted above, as: SAC = NOA-LCE > LCE = LCS.

The figures depicting the regression lines for analyses of covariance for the Vocabulary, Capitalization, Punctuation, Map Reading, Knowledge and Use of Reference Materials, Work-Study Skills--Total and Composite subscales seem to indicate that the LCE and LCS treatments were more effective for children with lower IQ's while children with higher IQ's performed better in treatments containing a limited amount of oral language instruction.

After the analyses of variance and covariance were performed, one additional comparison was made. For each subscale of the Iowa Tests of Basic Skills the mean score of the

third grade LCE group was compared to that of the fourth grade LCE group and the difference computed. The same thing was done for the LCS and SAC groups. Table 12 lists the mean scores for each treatment and the differences between them.

The interpretation of the comparisons of the third and fourth grade scores is based on the assumption that the only difference between the children in the sample of the two grades is one additional year in school. In other words, they need to have begun first grade with about the same characteristics and abilities and to have had about the same experiences since first grade in order for the comparisons to be valid. Data were not available to verify the assumption. It did happen that the IQ scores for the two grade levels in each treatment were very similar.

The scores shown on Table 12 are grade equivalent scores. The numeral before the decimal represents years and the one after the decimal represents months. Thus, as the table shows, the difference between the third and fourth grade mean scores for the LCE treatment ranged between seven months for Vocabulary and one year and five months for Spelling. The scores for all the tests except Spelling display differences of at least nine months, which is the normal length of the school year. If these differences can be interpreted as a reflection of growth, then the growth is quite satisfactory.

Table 12

Comparison of Fourth and Third Grade Mean Scores of Each Treatment Group on Iowa Tests of Basic Skills for Year Four (1967-68), San Antonio

	ICE			LCS			SAC		
	Third	Fourth	Difference	Third	Fourth	Difference	Third	Fourth	Difference
I.Q.[1]	88.88	88.72		82.30	87.02		91.96	91.62	
Vocabulary	2.7[2]	3.4	.7	2.5	3.3	.8	3.2	3.5	.3
Reading	2.9	3.8	.9	2.6	3.5	.9	3.3	3.8	.5
Spelling	3.1	4.6	1.5	2.8	3.7	.9	3.8	4.0	.2
Capitalization	3.0	4.0	1.0	2.6	3.7	1.1	3.8	3.8	.0
Punctuation	3.3	4.5	1.2	2.9	3.8	.9	4.0	4.0	.0
Usage	3.0	4.5	1.5	2.6	3.8	1.2	3.8	3.8	.0
Total Language	3.1	4.4	1.3	2.7	3.8	1.1	3.8	3.9	.1
Maps	2.9	4.2	1.3	3.0	3.7	.7	3.5	3.9	.4
Graphs	3.0	4.2	1.2	2.8	3.7	.9	3.4	4.0	.6
References	3.1	4.1	1.0	2.9	3.7	.8	3.4	4.0	.6
Total Work-Study	3.0	4.2	1.2	2.9	3.7	.8	3.5	4.0	.5
Arith. Concepts	3.2	4.1	.9	3.0	3.6	.6	3.6	3.9	.3
Arith. Problems	3.3	4.2	.9	2.9	3.6	.7	3.6	3.8	.2
Total Arithmetic	3.2	4.1	.9	2.9	3.6	.7	3.6	3.8	.2
Composite	3.0	4.0	1.0	2.7	3.6	.9	3.5	3.8	.3

[1] I.Q. as measured by The Lorge-Thorndike Intelligence Tests.

[2] All scores on Iowa Tests of Basic Skills are grade equivalents, i.e., 2.7 = second grade, seventh month.

The differences between the third and fourth grade mean scores for the LCS treatment ranged from six months for Arithmetic Concepts to one year and two months for Usage. The differences overall appear to be a little less than those of the LCE treatment, but nevertheless they reflect fairly adequate growth.

The differences for the LCE and LCS treatments are in striking contrast to those of the SAC group. These differences range between zero months (for three tests) and six months (for two tests).

Upon analyzing data derived from her language test, Taylor found negative differences in Fluency and Total scores between the fourth grade and fifth grade control groups in Year Five but positive differences for the LCE and LCS groups.[1] The children in Taylor's study in the LCE and LCS groups were the same children as those in this study minus the ones transferred to non-experimental classes. Her control groups consisted of different children who came from a school not in the Project but with a population very similar to that of the Project schools.

Taylor's findings with regard to language in conjunction with the ones in this study offer some evidence for the following hypothesis: The English language skills of native Spanish-speaking children in classrooms where no

---

[1]Taylor, "Year Five Findings," p. 110.

intensive language training is offered will develop to a point and then reach a plateau. This plateau in language development then limits growth in academic skills such as those measured by the Iowa Tests of Basic Skills.

CHAPTER IV

SUMMARY, LIMITATIONS, CONCLUSIONS, DISCUSSION AND RECOMMENDATIONS

SUMMARY

This study attempted to evaluate the effects on pupil performance of three and four years of intensive oral language instruction in English or Spanish. The subjects were disadvantaged third and fourth grade Mexican-American children who had been members of experimental classes since first grade in the San Antonio Language Research Project. The original goal of the Project was to enhance reading readiness. As the experimental treatments extended into higher grades the objective changed to developing skills in reading and in other academic subjects.

Originally there were three treatments. Two treatments included intensive language instruction, one in English (OAE--oral-aural English) and one in Spanish (OAS--oral-aural Spanish), and used science exercises from the AAAS Science: A Process Approach as the content base. Later, these treatments were referred to as the LCE (Language-Cognition English) and LCS (Language-Cognition Spanish) groups. Children in the initial third treatment, known as NOA (non-oral-aural), participated in the science activities but received no intensive language instruction. After the third year (1966-67) the NOA treatment, which had been considered the "control"

group, was dropped by the Project administrators. These children were then assigned, insofar as possible, to the English and Spanish experimental classes.

The attempt was made in this study to include in the data analysis scores only from children who had been in the same type of experimental treatment for all their years in school. The one exception made was to include children in the experimental English classes in third grade who had been in NOA classes during the first and second grades. This group was referred to as the NOA-LCE group in this study.

Because the original NOA treatment had been eliminated, another "control" group was selected, which consisted of children in the same schools but who had never been in experimental classes. They received only the regular San Antonio curriculum (not defined) and thus were referred to as the SAC group. It was assumed that any differences that might occur between the scores of the "control" group and the experimental groups could be attributed to the experimental curriculum because the children in the Project schools were to have been assigned randomly to experimental or non-experimental classes when they entered first grade.

The table below shows the number of children included in the research sample:

Treatment Group	Grade Four	Grade Three
LCE (initially OAE)	29	109
NOA-LCE	--	59
LCS (initially OAS)	60	40
SAC	21	68

For the purposes of this study it was assumed that the children receiving special instruction in oral language had improved in language ability. The question became one of investigating the effect of improved language abilities on reading and abilities in other academic subject areas. The instrument used to measure these abilities was the Iowa Tests of Basic Skills which is a written group test measuring skills in vocabulary, reading comprehension, language, work-study and arithmetic. The non-verbal battery of The Lorge-Thorndike Intelligence Tests was used to estimate differences in learning capacity.

Analysis of variance was used to compare the mean scores from the spring administration of the Iowa Tests of Basic Skills for the three treatment groups at the fourth grade level and the four treatment groups at the third grade level. Analysis of covariance also was used to compare the mean scores. The purpose of this analysis was to control for IQ, as measured by The Lorge-Thorndike Intelligence Tests administered to children in both grades in the fall of the year they were in third grade.

Findings

Fourth Grade

According to analysis of variance (see Table 8), there were significant differences among the three treatment groups only on Language Skills--Spelling ($p < .004$), Work-Study Skills--Total ($p < .048$), Arithmetic Skills--Arithmetic Concepts ($p < .046$) and Arithmetic Skills--Total ($p < .046$). In all cases, the LCE group had the highest mean score and the LCS group the lowest mean score. This generalization, based on analysis of variance, describes the differences which tended to occur: LCE > SAC > LCS.

When analysis of covariance with IQ as the covariable was applied, the differences listed above disappeared and two others occurred. Significant differences were found in Language Skills--Usage ($p < .044$) and Language Skills--Total ($p < .045$). Comparisons of adjusted means on these two scales revealed that the LCE mean scores were highest, the SAC mean scores the lowest and the LCS mean scores close to but higher than the SAC scores. Therefore, this generalization is appropriate: LCE > LCS $\cong$ SAC.

Examination of three figures (1, 2 and 5) displaying the results of analysis of covariance indicated that children with lower IQ's, as measured by The Lorge-Thorndike Intelligence Tests, benefitted from being in the LCE or LCS treatments while children with higher IQ's performed better if they were in the SAC group.

Table 13 provides a summary of the relationships of the mean scores of the treatment groups in the fourth grade on the Iowa Tests of Basic Skills according to analysis of variance and analysis of covariance. No scores are reported as this was done previously in Tables 8 and 9. Table 13 lists the groups according to the order of their means. Therefore, if a relationship reads LCE = SAC = LCS, this means that according to analysis of variance or of covariance there were no significant differences, but the LCE group mean score was highest, the SAC second, and the LCS group mean lowest. If a relationship includes only equal signs, that means that, according to analysis of variance or of covariance, there were no significant differences among the groups. In every case where there were significant differences among the groups according to analysis of variance, t-tests were applied to each pair of groups. Inequality signs in the middle column indicate the results of the application of t-tests. In the right column inequality signs were obtained by visual examination of the figures displaying the regression lines taken from analysis of covariance in the cases where significant differences had been found.

Third Grade

According to analysis of variance (see Table 9), there were significant differences on every test of the Iowa Tests of Basic Skills. The SAC group always obtained the highest mean scores. The LCS group received the lowest

Table 13

Summary of Results of Analysis of Variance and Analysis of Covariance, San Antonio Language Research Project, Grade Four, Year Four (1967-68)

Iowa Tests of Basic Skills	Analysis of Variance[1]	Analysis of Covariance with IQ as Covariable[2]
Vocabulary	SAC=LCE=LCS[3]	Analysis inappropriate
Reading	SAC=LCE=LCS	LCE=SAC=LCS
Spelling	LCE=SAC=LCS (LCE> LCS)	Analysis inappropriate
Capitalization	LCE=SAC=LCS	LCE=LCS=SAC
Punctuation	LCE=SAC=LCS	LCE=LCS=SAC
Usage	LCE=SAC=LCS	LCE > LCS > SAC
Total Language	LCE=SAC=LCS	LCE > LCS=SAC
Maps	LCE=SAC=LCS	LCE=SAC=LCS
Graphs	LCE=SAC=LCS	LCE=SAC=LCS
References	LCE=SAC=LCS	LCE=SAC=LCS
Total Work-Study	LCE=SAC=LCS (LCE > LCS)	LCE=SAC=LCS
Arith. Con.	LCE=SAC=LCS (LCE> LCS)	LCE=SAC=LCS
Arith. Prob.	LCE=SAC=LCS	LCE=SAC=LCS
Total Arith.	LCE=SAC=LCS (LCE> LCS)	LCE=SAC=LCS
Composite	LCE=SAC=LCS	Analysis inappropriate

[1]Equality and inequality signs reflect application of t-tests ($p < .05$) as well as analysis of variance. Equal signs indicate that differences in mean scores were not statistically significant.

[2]Equality and inequality signs reflect visual examination of Figures 3 and 4 as well as analysis of covariance.

[3]The groups are listed according to the order of their mean scores, with the highest first.

mean score on every test except the Map Reading subtest. The mean scores of the LCE and NOA-LCE groups were in the middle with the one exception noted before, with neither group being consistently superior to the other. Based on analysis of variance, this generalization describes the differences between the four groups: SAC > LCE = NOA-LCE > LCS.

Analysis of covariance was not always appropriate because the assumption that the regression lines be parallel was not met in every case. Of the eight instances when analysis of covariance was performed, significant differences were found in six: Reading Comprehension ($p < .014$), Language Skills--Spelling ($p < .001$), Language Skills--Usage ($p < .001$), Language Skills--Total ($p < .001$), Work-Study Skills--Reading Graphs and Tables ($p < .028$), and Arithmetic Skills--Total ($p < .030$). Examination of the adjusted mean scores of the four treatment groups on these six tests indicated that a fair generalization would be: SAC = NOA-LCE > LCE = LCS.

Observation of six of the figures (6, 9, 10, 13, 15, 16 and 18) displaying the results of analysis of covariance indicated that children with lower IQ's benefitted from being in the LCE or LCS treatments while children with higher IQ's performed better if they were in the SAC or NOA-LCE groups.

Table 14 provides a summary of the mean scores of the treatment groups in the third grade on the Iowa Tests

Table 14

Summary of Results of Analysis of Variance and Analysis of Covariance, San Antonio Language Research Project, Grade Three, Year Four (1967-68)

Basic Skills	Analysis of Variance[1]	Analysis of Covariance with IQ as Covariable[2]
Vocabulary	SAC > NOA-LCE=LCE=LCS[3] (NOA-LCE > LCS)	Analysis inappropriate
Reading	SAC > NOA-LCE=LCE > LCS	SAC=NOA-LCE > LCE=LCS
Spelling	SAC > NOA-LCE=LCE=LCS	SAC=NOA-LCE > LCS=LCE
Capitalization	SAC > LCE=NOA-LCE=LCS (LCE > LCS)	Analysis inappropriate
Punctuation	SAC > LCE=NOA-LCE=LCS (LCE > LCS)	Analysis inappropriate
Usage	SAC > LCE > NOA-LCE=LCS	SAC > LCE=NOA-LCE=LCS
Total Language	SAC > LCE=NOA-LCE=LCS (LCE > LCS)	SAC > NOA-LCE > LCE=LCS
Maps	SAC > LCS=LCE=NOA-LCE	Analysis inappropriate
Graphs	SAC > NOA-LCE=LCE=LCS	SAC=NOA-LCE > LCS=LCE
References	SAC > LCE=NOA-LCE=LCS	Analysis inappropriate
Total Work-Study	SAC > LCE=NOA-LCE=LCS	Analysis inappropriate
Arith. Con.	SAC > LCE=NOA-LCE=LCS	SAC=NOA-LCE=LCS=LCE
Arith. Prob.	SAC=NOA-LCE=LCE > LCS	NOA-LCE=SAC=LCE=LCS
Total Arithmetic	SAC > LCE=NOA-LCE=LCS (LCE > LCS)	SAC=NOA-LCE > LCE=LCS
Composite	SAC > NOA-LCE=LCE > LCS	Analysis inappropriate

[1]Equality and inequality signs reflect application of t-tests ($p < .05$) as well as analysis of variance. Equal signs indicate that differences in mean scores were not statistically significant.

[2]Equality and inequality signs reflect visual examination of Figures 7, 8, 11, 12, 14 and 17 as well as analysis of covariance.

[3]The groups are listed according to the order of their mean scores, with the highest first.

of Basic Skills according to analysis of variance and analysis of covariance. The format of this table is similar to the one for Table 13 which is explained on page 213.

Long-term Patterns

In order to place the findings of the Year Four (1967-68) study in perspective, summaries for the fourth grade (Sample I) and third grade (Sample II) subjects are outlined in Tables 15 and 16. The findings related to the "control" group in Year Two (1965-66) are omitted because the selection of the members of that group violated the usual research design procedures, i.e., the control pupils were not similar to the experimental pupils in ethnicity, social class, academic achievement and language ability. (However, it should be noted that the inclusion of that "control" group in the Year Two (1965-66) study did reveal the standing of the experimental groups in relation to the total community.) The nomenclature of the Year Four study is used, i.e., intensive English treatment = LCE and intensive Spanish treatment = LCS.

Sample I

The summary of findings related to the fourth grade research sample reveals that:

(1) At the end of first grade, there were no differences among the groups on reading readiness;

Table 15

Summary of Findings Related to Sample I (children who entered first grade in 1964) in the San Antonio Language Research Project at the End of Year Four (1967-68)

First Grade (Horn, "Year One Findings")

Criterion: Reading Readiness

NOA = LCE = LCS

Second Grade (Arnold, "Year Two Findings")

Criterion: Reading

NOA > LCE = LCS

Third Grade (Knight, "Year Three Findings")

Criterion: Reading

nine significant differences--

NOA > LCE > LCS

six significant differences--

LCE > NOA > LCS

Fourth Grade (Salinas, "Year Four Findings")
NOA group discontinued

Criterion: Iowa Tests of Basic Skills

analysis of variance--

LCE > SAC > LCS

analysis of covariance with IQ as covariable

LCE > LCS ≅ SAC

NOA = Non-Oral-Aural
LCE = Language-Cognition English
LCS = Language-Cognition Spanish
SAC = San Antonio Curriculum (control)

Table 16

Summary of Findings Related to Sample II (children who entered first grade in 1965) in the San Antonio Language Research Project at the End of Year Four (1967-68)

First Grade (Arnold, "Year Two Findings")

Criterion: Reading

LCE > NOA = LCS

Criterion: Reading Readiness

high pretest scores--

LCE = NOA = LCS

low pretest scores--

LCE = NOA > LCS

Second Grade (Knight, "Year Three Findings")

Criterion: Reading

LCE > NOA > LCS

Third Grade (Salinas, "Year Four Findings")

Criterion: Iowa Tests of Basic Skills

analysis of variance--

SAC > LCE = NOA-LCE > LCS

analysis of covariance with IQ as covariable--

SAC = NOA-LCE > LCE = LCS

NOA = Non-Oral-Aural
LCE = Language-Cognition English
LCS = Language-Cognition Spanish
NOA-LCE = NOA for two years, LCE in third grade
SAC = San Antonio Curriculum (control)

(2) At the end of second grade, the NOA group was superior to the LCE and LCS groups in reading while the LCE and LCS groups were about the same;

(3) At the end of third grade, nine significant differences on a reading measure favored the NOA group and six favored the LCE group. The LCS group was always third in order; and

(4) At the end of fourth grade the LCE group had a slight advantage on the Iowa Tests of Basic Skills whether or not IQ was controlled. When IQ was not considered, the control (SAC) group had higher scores than the LCS group. When IQ was considered, the LCS group scored slightly higher or about the same as the control (SAC) group.

In Sample I a trend appeared to be developing up through third grade which favored the NOA treatment. Unfortunately, the NOA treatment was discontinued after third grade so there is no way of knowing for sure whether or not the LCS group would have scored highest on the Iowa Tests of Basic Skills if the NOA group had been included in the comparison. Although the NOA pupils received special instruction with science materials, they probably received approximately the same kind of oral language instruction as the "control" pupils, e.g., no planned formal language instruction at all. Therefore, it is assumed here that the NOA group of the first three years was similar to the "control" group of the fourth year. It follows then that

the trend favoring the groups not receiving the experimental language curricula was reversed in fourth grade and the LCE treatment proved slightly more effective.

Sample II

The summary of findings related to the third grade research sample in Year Four shows that:

(1) At the end of first grade, when reading readiness was measured, the LCE, LCS and NOA groups were about the same except when scores of pupils with low pretest scores were compared. In this case, the LCS group scored lower than the other two groups. When reading skills were measured, the LCE group was superior to the NOA and LCS groups, which were about the same.

(2) At the end of second grade, tests of reading showed that the LCE group was superior to the NOA group, which was in turn superior to the LCS group.

(3) At the end of third grade, the "control" group scored highest on the Iowa Tests of Basic Skills. There were no consistent differences between the LCE and NOA-LCE groups. The LCS group scored lowest. When IQ was controlled, the NOA-LCE scores were found to be greater than those of the LCE group and about the same as the SAC scores. The LCS scores were about the same as the LCE scores.

For the first two years, the LCE treatment appeared to be slightly more effective than other treatments for the children in Sample II. The test scores in the third grade

did not provide evidence that the LCE treatment was more effective than the language activities (not defined) in the regular San Antonio curriculum. Thus, for the first time, pupils in the intensive language programs who had teachers who were in their second year of experience with the new program, did not out-perform pupils receiving no special language instruction.

LIMITATIONS

Instrumentation

Performance on the Iowa Tests of Basic Skills depends heavily on the ability to read. On every subtest, even the arithmetic sections, the child has to read the questions before he can display the skills that are supposed to be assessed in that section of the test. The mean scores of the fourth grade groups ranged between 3.5 (third grade, fifth month) and 3.8. Those of the third grade experimental groups ranged between 2.6 and 2.9, while the SAC group mean was 3.3. These low reading levels indicate that the children in the research sample were likely to have been unduly handicapped by their reading abilities on the tests assessing skills in other areas.

Adding to the problem is the fact that both the children and their teachers, when faced with a test that appears to them to be very difficult, do not seem to take the testing sessions seriously. Thus the children in the

research samples may not have performed as well as they might have if they had had more confidence in themselves and in the process of taking exams. It may be true, however, that these factors relating to the difficulty of the instrumentation used in this study affect all treatment groups about the same.

Sample Size

By the fourth year (1967-68) of the Project, the sample sizes were reduced to as few as twenty-nine in the fourth grade LCE treatment group compared to the 250 children in each treatment group in 1964. The number of children in experimental classes was still large, but the number who had received the same treatment for three or four years was much smaller. Unfortunately there was no way of knowing whether the children in the research sample in Year Four fairly represented the spectrum of abilities of the original sample. It is impossible to say whether or not the effects of the intensive language instruction were the same on those children who advanced to the third or fourth grade in the same experimental treatment as the effects would have been on all the children who started receiving experimental language instruction in first grade.

Lack of Monitoring

Because of lack of resources, the teachers were not supervised on a daily basis either by the Project staff or by the supervisory staff of the San Antonio Independent

School District. The teachers were not asked to report the amount of time they spent teaching the experimental curriculum but were assumed to follow project directions. Therefore, even though the experimental design called for instruction in the special curriculum for an hour a day for 140 days a year, no one knows if these instructions were actually carried out. Observers suspect that there was great variation in use of the experimental curriculum from classroom to classroom.

Another source of variation occurred in the degree to which the teachers used the oral-aural techniques in lessons during the day. It is likely, though, that teachers in the Project for the first or second year would not have felt enough confidence to try developing their own language drills for use the rest of the day. This source of variation might be more important in later years of research.

Some deviation from the planned instruction occurred because production of materials was delayed. This is known to have most strongly affected the implementation of the LCS treatment and the fourth grade LCE classes the year they were in third grade.

Nature of the San Antonio Curriculum

The comparisons made in the research in the Project generally referred to the regular San Antonio curriculum. For instance, one of the three methods of developing reading readiness in the Year One (1964-65) study was whatever the

San Antonio Independent School District stipulated. In later years the intensive language programs were compared to whatever methods were used by the teachers in the non-experimental classes. Unfortunately, no one knows specifically what methods the teachers in non-experimental classes used or even what methods were used by the teachers in the LCE and LCS classes during the part of the day not involving the language/science materials. It seems safe to assume that although teachers' methods and styles varied somewhat, audiolingual language teaching methods, with few exceptions, were not used outside of Project classrooms and probably little attention was paid to structured language development per se. A few of the NOA teachers may have occasionally used audiolingual techniques but they did not do so on a regular systematic basis. Probably none of the "control" teachers used them.

Assessment of Oral Language

The objective of the San Antonio Language Research Project was to improve oral language abilities in order to enhance academic achievement of Spanish-speaking children. Because oral language assessment instruments which have general acceptance were not available, no firm conclusions could be reached about the effects of the different language teaching methods used. The use of reading readiness tests, reading tests and the Iowa Tests of Basic Skills was predicated on the assumption that the language of the children

in the experimental classes had improved. The data from the written tests would show whether or not the improved language abilities would affect other kinds of school performance. The problem with this assumption and the lack of data on oral language during every year of the Project is that if significant differences are not found on the reading and other tests, then one does not know whether the experimental curriculum failed to develop language or whether improved language has no effect on other skills.

Assignment to Treatment Groups

When the Project was begun the school administrators agreed to assign pupils to treatment groups on a random basis. There is evidence that in a number of cases, pupils of lower age and less developed English abilities were assigned to the LCE classes. To the extent that the statistical techniques used in the various studies did not account for these initial differences, the effects of the intensive English instruction may have been underestimated.

Retention Policy

The failure rate of first grade Spanish-speaking children had been very high in some schools in San Antonio. Principals of Project schools agreed to promote all children in the experimental classes. In a very few cases this was not done. Data were eliminated in this study for any child

in the third grade LCE, LCS and NOA-LCE classes who had been retained.

A more difficult problem arose in that it is not known whether or not the schools applied the new policy of no failures to the non-experimental classes. If the teachers maintained a policy of failing the lowest achieving children only in the SAC (control) classes, these classes would have children not comparable to those in the experimental classes. This lack of similarity might be reduced by the extent to which the retainees of the next higher grade were members of the research sample classes.

Examination of the chronological ages of the children in the "control" classes indicated that at least one of the twenty-one fourth grade SAC pupils was overage for his grade and at least eight of the sixty-eight third grade SAC pupils were overage. There are other reasons besides failure for being overage but, whatever the explanation, these overage children probably represent the lower spectrum of abilities or achievement levels.

The interpretation of the comparisons of achievement of experimental and control pupils in this study is therefore limited by the lack of exact information on the composition of the control classes.

Teacher Variables

By the fourth year of the Project, the several children in each treatment in the research samples had received

instruction from many different teachers. It is hoped that this fact will mean that teacher variation did not unduly affect the data.

In addition to the usual differences in teacher competency, other factors related to the Project contributed to teacher differences. Success which pupils would have had with the Spanish lessons depended in part on their teachers' mastery of Spanish. The teachers who participated in the Project had to spend much more time in preparation and in inservice meetings than non-Project teachers. This extra burden may have affected the teachers' attitudes and willingness to cooperate in varying degrees. The principals of the various buildings also differed in their willingness to cooperate. Finally, the Project staff supervisors apparently varied a great deal in the performance of their duties due in part to different philosophical points of view and in part to the erratic observation schedules.

CONCLUSIONS

Within the limitations set forth above, the data in this study allow these conclusions to be made about the effects on children in San Antonio Language Research Project schools at the conclusion of the Project's fourth year (1967-68) in the LCE, NOA-LCE and LCS treatments:

(1) The LCS (Language-Cognition Spanish) treatment was not effective in producing differential achievement as measured by the Iowa Tests of Basic Skills.

(2) According to data from the fourth grade sample, the LCE (Language-Cognition English) treatment was slightly more effective in producing differential achievement than language activities of the regular San Antonio curriculum. Significant differences favoring the LCE treatment were found in comparisons of mean scores on the following tests of the Iowa Tests of Basic Skills: Language Skills--Spelling; Work-Study Skills--Total; Arithmetic Skills--Arithmetic Concepts; and Arithmetic Skills--Total. When IQ was controlled the significant differences favoring the LCE treatment were found in Language Skills--Usage; and Language Skills--Total.

(3) According to data from the third grade sample, the SAC (San Antonio curriculum) was more effective in producing differential achievement as measured by the Iowa Tests of Basic Skills. When IQ was controlled, the NOA-LCE (Non-Oral-Aural/Language-Cognition English) treatment was also more effective than the experimental language treatments--LCE and LCS.

(4) The superiority of the SAC treatment at only the third grade level may be an anomaly or it may represent a pattern.

(5) On the subscales where analysis of covariance was not appropriate because of interaction effects, figures displaying the regression lines revealed a tendency for the SAC and NOA-LCE treatments to be more effective at the

higher IQ levels and for the LCE and LCS treatments to be more effective at the lower IQ levels.

(6) The grade equivalent scores and the percentile rankings of the children on the Iowa Tests of Basic Skills in all the treatments--experimental as well as control--were very low. At the end of fourth grade the Composite mean scores of the Sample I children ranged between 3.6 (third grade, sixth month) and 4.0 (beginning of fourth grade). At the end of third grade the Composite mean scores of the Sample II children ranged between 2.7 (second grade, seventh month) and 3.5 (third grade, fifth month). The percentile rankings ranged between the fourth and thirty-ninth percentiles for children in fourth grade and between the third and sixty-second percentiles for children in the third grade.

DISCUSSION

Spanish Treatment Findings

The most consistent set of findings in the Project research related to the ineffectiveness of the LCS treatment. On almost no measure did the LCS groups receive scores higher than the other experimental or control groups. These results, however, should not be interpreted as an indictment against the use of Spanish in the classroom or against bilingual education. The major instruments used for assessment, both oral and written, were in English, not Spanish. The data from the administration of the Spanish reading tests in Year Two

(1965-66) and Year Three (1966-67) were not favorable to the LCS treatment, but instruction in reading in Spanish was not a component of the experimental or the regular curriculum.

In addition to the lack of proper testing, certain shortcomings were seen in the Spanish materials which must have been factors in the effectiveness of the treatment. The Spanish lessons consisted of translations of the English lessons. Thus, they were not based, even at first and second grade levels, on any systematic outline of the syntactic structures of Spanish or on any other logical plan. Moreover, the development of the materials was not preceded by any diagnosis of the language abilities of the pupils. It is also known that the Spanish curriculum materials frequently arrived in San Antonio late so that the teachers were not able to teach new material in the daily time allotted to their experimental treatment. This lack of curriculum materials was especially serious for the Spanish-surname teachers who were not fluent in Spanish (the Project administrators had erroneously assumed that all Spanish-surname teachers spoke Spanish). Therefore, the negative findings in this study with regard to the Spanish treatments are not surprising.

English Treatment Findings

It was assumed that if the experimental English curriculum were effective, there would be significant differences

favoring the LCE group on scores of the Iowa Tests of Basic Skills. Improved skills in language would be reflected in higher scores on the Vocabulary, Usage and perhaps the Spelling tests. Improved language would be reflected also in higher Reading Comprehension scores. Because the content used by all three treatments (LCE, LCS and NOA) included lessons with science and math concepts and processes, the experimental treatment groups would be expected to score higher than the "control" groups on both arithmetic tests and on the Graphs and Map Reading subtests. Finally, if the experimental language curricula were effective in producing growth in English language, the LCE and LCS groups would be expected to score higher on all tests because their improved language skills would enable them first, to read the questions better and second, to be more likely to know the answers because they had understood their teachers better.

All the expected differences did not occur. This does not mean that intensive language, or language activities in combination with AAAS lessons, are not effective. The findings of this study indicate only that the materials were not as effective as was hoped for in the San Antonio Language Research Project during the years 1964 to 1968. No other conclusion can be fairly made.

San Antonio Curriculum Findings

There is no obvious reason why the SAC group was superior at the third grade level but not at the fourth

grade level. The results from the two grade levels may be an unexplained contradiction. On the other hand, it is possible that children learn a great deal of language by being exposed to it in instructional settings in traditional classrooms. However, at some point, without intensive language training, their growth in language reaches a plateau. This may have happened to the SAC children in the fourth grade. This hypothesis is supported by Taylor's findings that fifth grade "control" children scored lower than fourth grade "control" children in Fluency and Total Language on her test (Taylor's fifth grade and fourth grade "control" pupils came from the same population as the fourth and third grade pupils, respectively, in this study). An alternative explanation is that the effects of intensive language instruction may be long run in nature and do not necessarily reveal themselves until after four years of continuous instruction.

Since no oral language assessment of the pupils was made during Year Four (1967-68), no conclusion can be reached as to whether the LCE and LCS treatments were not effective in improving academic achievement even though language growth occurred, or whether the experimental language treatments produced differential achievement in neither language nor academic areas.

NOA-LCE Treatment Findings

The finding that the scores on the Iowa Tests of Basic Skills of the LCE and NOA-LCE treatment groups were very

similar according to analysis of variance may be interpreted in at least three ways. (All three assume that improved language abilities would be reflected in scores of written tests in various curriculum areas.)

(1) The NOA treatment (which the third grade pupils had received for two years) was not very different from the LCE treatment. This may have been because the NOA teachers used the oral-aural techniques they learned about by talking to LCE teachers. If this is true, it implies that the extensive inservice sessions the LCE teachers attended may not be any more effective for developing teaching techniques than informal teacher-to-teacher contacts.

(2) The LCE treatment was not any more effective in developing oral language abilities than the informal language activities that were carried on by NOA teachers and presumably by teachers of non-experimental classes.

(3) The NOA treatment was just as effective as the LCE treatment because of a specific characteristic of the approach to teaching science. The amount of teacher talk is generally much smaller than in the lecture mode typically used in other subject areas. The pupils talk to the teacher more than usual in large group discussions and to each other in small group discussions as they work on science project.

Influence of IQ

There was a tendency at both grade levels for the children with higher IQ's to perform better than other treatment

group members on the Iowa Tests of Basic Skills if they were in the SAC treatment group, while children with lower IQ's tended to perform better if they were in the experimental language treatments (LCE and LCS) than if they were in the SAC treatment. This result was not anticipated by the San Antonio Language Research Project staff members. There is no way of knowing for sure which interpretation of the data would be most accurate. However, at least one hypothesis appears plausible at this time.

If child performance is systematically handicapped on an IQ test because of limited language ability, then it would appear that it is possible that children with little facility in English were helped by the experimental English or Spanish curricula. Children with little control of English, and therefore lower measured IQ's, who happened to be assigned to SAC treatment classes, were not given planned and prolonged instruction in language. Therefore, they were handicapped in their development of the various skills tapped by the Iowa Tests of Basic Skills. However, if these children were in the LCE or LCS classes, they were given the kind of help in language they needed and thus were able to perform better on the Iowa Tests of Basic Skills.

Children who had managed to learn a great deal of English as indicated by superior performance on the IQ tests and received the SAC treatment proceeded to develop their academic skills with no linguistic handicap. On the other

hand, children with well developed English language skills in the LCE and LCS treatment groups were actually held back in comparison to their peers in the SAC treatments, because their teachers were spending time providing language lessons that these children did not need.

This interpretation implies a need for teachers to tailor their oral language instruction to the specific needs of their individual children. Approaches such as the one used in San Antonio in which language lessons are taught to entire classes with no grouping or individualization are inefficient and may even prove harmful to children whose needs are not being met by the large group lesson.

The need for individualization of oral language instruction means that there is an even greater necessity for two things:

(1) Development and refinement of instruments for assessing oral language which teachers can use as diagnostic pretests.

(2) Development of methods for training teachers how to assess oral language informally and how to individualize instruction in general in the self-contained classroom.

Overall Results

Interpretations of the favorable results for the SAC treatment at the third grade level and the LCE treatment at the fourth grade level should be made in view of the fact that the scores of the children in all the treatments,

experimental and control, at both grade levels were very low. Both McDowell in his study of first graders in Year One (1964-65) and Arnold in his study of first graders in Year Two (1965-66) found that the children in the Project schools performed very poorly on a variety of intelligence, reading and reading readiness tests in comparison to other children in the community who came from diverse socioeconomic classes and ethnic and language groups. There is a clear need for a fresh look at the problems of the children in the Project schools, and all children like them, and a necessity for intensive efforts to find ways to improve their instruction including, perhaps, a regrouping of the entire educational structure.

The many difficulties these children face in school obviously are diverse and severe. The San Antonio Language Research Project tried to identify the problems and attempted to alleviate many of them. That the attempt was not successful is an indication of the difficulty and complexity of the situation.

Worth of the Project

Since the experimental treatments did not produce great improvements in skills in reading and other academic areas, an appropriate question is whether or not the Project was worth the time, money and effort invested in it. The opinion of this writer is that the San Antonio Language Research Project did make contributions in at least the following ways:

(1) Data which were gathered established dramatically the inappropriateness for linguistically different students of a number of reading, reading readiness and intelligence tests and the lack of adequate oral language assessment instruments.

(2) By highlighting the problems to be faced in doing research in public schools, the San Antonio Language Research Project provided information to other groups about what they should strive to achieve, avoid and/or overcome.

(3) The Project "ploughed ground" for language programs which followed and are continuing and which focus upon problems of linguistically different learners.

RECOMMENDATIONS

The findings of this study and the experience gained as a researcher in the San Antonio Language Research Project prompt the proposal of these recommendations:

(1) Any group doing longitudinal methodological research in public schools should take care to:

(a) Obtain prior commitment of administrators and principals, which must be communicated to the staff.

(b) Assure that sufficient monitoring of teacher and pupil behavior be done so that data obtained in testing are meaningful and can be interpretated.

(c) Provide continuing staff development in the way of both consultant services and moral support.

(d) Assure that critical teacher feedback to the materials development staff is transmitted and considered and acted upon.

(e) Include as large a component as possible which explains to all personnel involved the research and development process, why evaluation of both teacher and pupil performance and attitude is needed, and why instructional monitoring is necessary.

(f) Secure services of a Project director who is likely to stay for the length of the project.

(2) Additional research with regard to oral language instruction and testing should be conducted in order to determine:

(a) What kinds of oral language instruction are most effective in improving oral language proficiency.

(b) What amount of oral language instruction is appropriate.

(c) If oral language instruction should be a separate component or a part of all areas of the curriculum.

(d) How oral language instruction should be related to other areas of instruction.

(e) What aspects of oral language are most closely related to success in other school endeavors.

(f) Whether instruction in other areas should be modified to meet the needs of non-native English speakers and, if so, how.

(g) If special materials need to be developed for oral language instruction, or if emphasis should be on staff development to train teachers to construct their own language drills and activities which can be used with whatever materials are available in the school district.

(3) Work should continue on the development and refinement of instruments and methods for assessing oral language facility. Progress in finding answers to the questions listed in the second recommendation hinges on the availability of such instruments and methods.

APPENDIX A

ADDITIONAL SOCIOECONOMIC DATA

Income

Considering the poverty line to be $3,000 income per family per year, the following percentage of families are "poor" in the Southwest as of 1960:

Anglo--15.9

Spanish-surname--34.8[1]

In Texas in 1960 over half of all Spanish-surname families had annual incomes under $3,000.[2]

Housing

Housing is considered overcrowded if there are 1.01 or more persons per room. In 1960 the following percentage of housing was "crowded":

Southwestern metropolitan areas:

Anglo--7.7

Spanish-surname--34.6

Texas: Anglo--9.4

Spanish-surname--46.3[3]

---

[1]Frank G. Mittlebach and Grace Marshall, "The Burden of Poverty," Mexican-American Study Project Advance Report 5, University of California, Los Angeles, 1966, pp. 2-3.

[2]Robert W. MacMillan, "A Study of the Effect of Socio-economic Factors on the School Achievement of Spanish-Speaking School Beginners," Unpublished Doctoral Dissertation, The University of Texas, Austin, 1966, p. 45.

[3]Mittlebach and Marshall, "The Burden of Poverty," p. 44.

Occupation

Sixty percent of all employed urban Mexican-American males were in three "manual" classifications according to the Census Bureau: operatives (semiskilled), craftsmen and laborers. Only nineteen percent are in white collar occupations: professional, managerial, clerical and sales. Almost half of all employed Anglos are white collar workers.[1]

Unemployment

The schools in the San Antonio Language Research Project are in an area which has the highest subemployment rate of ten areas studied by the U.S. Department of Labor, areas which included Harlem, East Harlem and Bedford-Stuyvesant.[2] Forty-seven percent of the persons in the San Antonio slum areas who are or should be workers are subemployed.[3]

---

[1]Walter Fogel, "Mexican-Americans in Southwest Labor Markets," Mexican-American Study Project Advance Report 10, University of California, Los Angeles, 1967, pp. 15-16.

[2]The subemployment index includes people who are: jobless and looking for work; working part time although full time work is wanted; heads of households under six who earn less than $60 a week though working full time; or not heads of households and earning less than $56 a week on a full time job.

[3]U.S. Department of Labor, Manpower Report of the President and A Report on Manpower Requirements, Resources, Utilization, and Training, Transmitted to the Congress, April, 1967, United States Department of Labor, Washington, D.C., p. 76.

APPENDIX B

Number of Classrooms in
San Antonio Language Research Project
1964-1968

Year	Sample	Grade	OAE Classes	OAS Classes	NOA Classes	"CONTROL" Classes	TOTAL Classes	GRAND TOTAL Classes
1 (1964-65)	1	1st	10	10	10	--	30	30
2 (1965-66)	1	2nd	12	10	11	12	45	
2 (1965-66)	2	1st	12	7	11	12	42	87
3 (1966-67)	1	3rd	8	7	10	12	37	
3 (1966-67)	2	2nd	10	8	9	12	39	76
			LCE	LCS		CONTROL (SAC)		
4 (1967-68)	1	4th	14	8		10 (Apprx)	32	
4 (1967-68)	2	3rd	18	9		10 (Apprx)	37	
			ESL	ESL-T	Bi-T	Bi		
4 (1967-68)	3	2nd	8	3	3	12	26	
4 (1967-68)	3	1st	14	6	6	7	31	126

OAE = Oral-Aural English
OAS = Oral-Aural Spanish
NOA = Non-Oral-Aural

LCE = Language-Cognition, English
LCS = Language-Cognition, Spanish
SAC = San Antonio Curriculum

Bi = Bilingual Teacher
Bi-T = Team, Bilingual Teachers
ESL = English as a Second Language Teacher
ESL-T = Team, English as a Second Language Teachers

BIBLIOGRAPHY

Books, Papers and Reports

Andersson, Theodore. "Bilingual Education: The American Experience," Paper presented at the Ontario Institute for Studies in Education Conference on Bilingual Education, Toronto, March 13, 1971. (Mimeographed) Citing Texas Education Agency, Division of Research, "Reports of Pupils in Texas Public Schools Having Spanish Surnames, 1955-56." Austin, Texas: Texas Education Agency, 1957.

Arnold, Richard D. "Capsule Description of San Antonio Language-Bilingual Research Project." Austin, Texas: The University of Texas (Mimeographed), 1967.

_____. "1965-66 (Year Two) Findings, San Antonio Language Research Project, Thomas D. Horn, Director." Austin, Texas: The University of Texas, 1968.

_____. "Retention in Reading of Disadvantaged Children," Reading and Realism, J. Allen Figurel, ed. Newark, Delaware: International Reading Assn., 1969.

Bogue, Donald J. The Population of the United States. Glencoe, Illinois: The Free Press, 1959.

Buros, Oscar K. (ed.). The Fifth Mental Measurements Yearbook. Highland Park, N.J.: The Gryphon Press, 1959.

Commission on Science Education, American Association for the Advancement of Science. Science: A Process Approach. Washington, D.C.: Commission on Science Education, American Association for the Advancement of Science, 1963.

Fogel, Walter. "Mexican-Americans in Southwest Labor Markets," Mexican-American Study Project Advance Report 10. Los Angeles: University of California, 1967.

Freeman, Frank S. Review of The Lorge-Thorndike Intelligence Tests, in The Fifth Mental Measurements Yearbook, Oscar K. Buros, ed. Highland Park, N.J.: The Gryphon Press, 1959.

Gloria and David Beginning English, Series No. 20, Test 6. Austin, Texas: Language Arts, Inc., 1958.

Gloria and David Spanish Intermediate, Series No. 60, "Saturday's Activities." Austin, Texas: Language Arts, Inc., 1958.

Governor's Committee on Public School Education. The Challenge and the Chance. Austin, Texas: Governor's Committee on Public School Education, 1968.

Horn, Thomas D. "A Study of the Effects of Intensive Oral-Aural English Language Instruction, Oral-Aural Spanish Language Instruction and Non-Oral-Aural Instruction on Reading Readiness in Grade One." Austin: The University of Texas, 1966.

Jennings, E. A Subroutine System for Data Processing in the Behavioral Sciences. Austin: The University of Texas, 1968.

Livermore, Arthur H. "The Process Approach of the AAAS Commission on Science Education," Readings in Science Education, Edward Victor and Marjorie S. Lerner, eds. New York: The MacMillan Company, 1967.

Manuel, Herschel T. Spanish-Speaking Children of the Southwest. Austin: The University of Texas Press, 1965.

McNemar, Quinn. Psychological Statistics, 3rd ed. New York: John Wiley and Sons, Inc., 1962.

Mittlebach, Frank G. and Marshall, Grace. "The Burden of Poverty," Mexican-American Study Project Advance Report 5. Los Angeles: University of California, 1966.

Morgan, G.A.V. Review of Iowa Tests of Basic Skills, in The Fifth Mental Measurements Yearbook, Oscar K. Buros, ed. Highland Park, N.J.: The Gryphon Press, 1959.

Pidgeon, D.A. Review of The Lorge-Thorndike Intelligence Tests, in The Fifth Mental Measurements Yearbook, Oscar K. Buros, ed. Highland Park, N.J.: The Gryphon Press, 1959.

Saunders, Lyle. Address given to the National Convention, League of United Latin American Citizens, San Antonio, Texas, June 11, 1949. (Mimeographed) Quoted by Carey McWilliams, Brothers Under the Skin. Boston: Little, Brown and Company, 1951.

Stockwell, Robert P.; Bowen, J. Donald; and Martin, John W. The Grammatical Structures of English and Spanish. Chicago: The University of Chicago Press, 1965.

Texas Education Agency. Pre-School Instructional Program for Non-English Speaking Children, Bulletin 642. Austin: Texas Education Agency, 1964.

_____. A Report on the Pre-School Instructional Program for Non-English Speaking Children. Austin: Texas Education Agency, 1962.

U.S. Bureau of the Census. Current Population Reports, Series P-20, No. 213, "Persons of Spanish Origin in the United States: November 1969." Washington, D.C.: Government Printing Office, 1971.

U.S. Department of Labor. Manpower Report of the President and A Report on Manpower Requirements, Resources, Utilization, and Training. Transmitted to the Congress, April, 1967. Washington, D.C.: United States Department of Labor, 1967.

Witty, Paul A. (ed.). "The Educationally Retarded and Disadvantaged," Sixty-Sixth Yearbook for the National Society for the Study of Education, Part I. Chicago: University of Chicago Press, 1967.

Theses and Dissertations

Bradley, N. Ruth. "A Study of the Relation of Oral Language Proficiency and Reading in a Group of Fourth Grade Children of a French Linguistic Background," Unpublished Master's Thesis, The University of Texas at Austin, 1970.

_____. "A Two-Year Investigation of the Relation of Oral Language Proficiency and Reading Achievement of First Grade Children with a French Linguistic Background," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1971.

Bussey, Jo Ann Keslar. "A Comparative Study of the Phonological Variations Among First, Second, and Third Grade Linguistically Different Children in Five San Antonio (Texas) Schools," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1971.

Cook, Marilyn H. "The Development of a Reading Test for an Experimental Reading Program for Disadvantaged Children," Unpublished Master's Thesis, The University of Texas at Austin, 1969.

Cornejo, Ricardo Jesus. "Bilingualism: A Study of the Lexicon of the Five-Year-Old Spanish-Speaking Children of Texas," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1969.

Jameson, Gloria R. "The Development of a Phonemic Analysis for an Oral English Proficiency Test for Spanish-Speaking School Beginners," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1967.

Knight, Lester N. "A Comparison of the Effectiveness of Intensive Oral-Aural English Instruction, Intensive Oral Spanish Instruction, and Non-Oral-Aural Instruction on the Reading Achievement of Spanish-Speaking Second- and Third-Grade Pupils," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1969.

MacMillan, Robert W. "A Study of the Effect of Socioeconomic Factors on the School Achievement of Spanish-Speaking School Beginners," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1966.

McDowell, Neil A. "A Study of the Academic Capabilities and Achievements of Three Ethnic Groups: Anglo, Negro, and Spanish-Surname, in San Antonio, Texas," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1966.

Melton, Jo Anne C. "A Comparison of Two Methods of Teaching Reading to Disadvantaged Mexican-American First Grade Children," Unpublished Master's Thesis, The University of Texas at Austin, 1968.

Ott, Elizabeth H. "A Study of Levels of Fluency and Proficiency in Oral English of Spanish-Speaking School Beginners," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1967.

Pauck, Frederick G. "An Evaluation of the SELF-Test as a Predictor of Reading Achievement of Spanish-Speaking First Grade Children," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1968.

Peña, Albar A. "A Comparative Study of Syntactical Structures of the Oral Language Status in Spanish and English of Disadvantaged First-Grade Spanish-Speaking Children," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1967.

Taylor, Thomasine H. "The Effects of Oral-Aural Language Training on Gains in English Language for Fourth and Fifth Grade Disadvantaged Mexican-American Children," Unpublished Doctoral Dissertation, The University of Texas at Austin, 1969.

Articles

Arnold, Richard D. "English as a Second Language," The Reading Teacher, XXI (April, 1968), 634-639.

Hillerich, Robert L. "An Interpretation of Research in Reading Readiness," Elementary English, XLIII (April, 1966), 359-364, 372.

Horn, Thomas D. "Three Methods of Developing Reading Readiness in Spanish-Speaking Children in First Grade," The Reading Teacher, XX (October, 1966), 38-42.

Stemmler, Anne O. "An Experimental Approach to the Teaching of Oral Language and Reading," Harvard Educational Review, XXXVI (Winter, 1966), 42-59.

Weintraub, Samuel. "Readiness Measures for Predicting Reading Achievement," The Reading Teacher, XX (March, 1967), 551-558.

Tests

Brengelman, Frederick H. and Manning, John C. Linguistic Capacity Index. Fresno: Fresno State College, 1964.

Cattell, R.B. Institute for Personality and Ability Testing (IPAT) Culture Fair Intelligence Test, Scale 1. Champaign, Illinois: Institute for Personality and Ability Testing, 1964.

Durost, Walter N. (ed.); Bixler, Harold N.; Hildreth, Gertrude H.; Lund, Kenneth W. Metropolitan Achievement Tests, Primary, Levels 1 and 2, Form A. New York: Harcourt, Brace and World, 1958.

_____. Metropolitan Achievement Tests, Primary I, Form A. New York: Harcourt, Brace and World, 1958.

Durost, Walter N. (ed.); Bixler, Harold N.; Hildreth, Gertrude H.; Lund, Kenneth W. Metropolitan Achievement Tests, Primary II, Form A. New York: Harcourt, Brace and World, 1958.

_____. Metropolitan Achievement Tests, Primary II, Form C. New York: Harcourt, Brace and World, 1961.

Durost, Walter N. (ed.); Bixler, Harold N.; Hildreth, Gertrude H.; Lund, Kenneth W.; and Wrightstone, J. Wayne. Metropolitan Achievement Tests, Primary I, Form C. New York: Harcourt, Brace and World, 1961.

Goodenough, F.L. and Harris, Dale B. Goodenough-Harris Draw-A-Man Test. New York: Harcourt, Brace and World, 1963.

Hildreth, Gertrude H.; Griffiths, Nellie L.; and McGauvran, Mary E. Metropolitan Readiness Tests, Form A. New York: Harcourt, Brace and World, 1964.

Lindquist, E.F. and Hieronymus, A.N. Iowa Tests of Basic Skills, Grade 4, Form 3 and Grade 3, Form 4. Boston: Houghton Mifflin Co., 1964.

_____. Manual for Administrators, Supervisors, and Counselors: Iowa Tests of Basic Skills. Boston: Houghton Mifflin Co., 1964.

Lorge, Irving and Thorndike, Robert L. The Lorge-Thorndike Intelligence Tests, Level 2, Form A, non-verbal battery. Boston: Houghton Mifflin Co., 1957.

Manuel, Herschel T. Prueba de Habilidad General, Serie Interamericana, Primario, Nivel 1, Forma CEs. Austin, Texas: Guidance Testing Associates, 1962.

_____. Prueba de Habilidad General, Serie Interamericana, Primario, Nivel 1, Forma DEs. Research edition. Austin, Texas: Guidance Testing Associates, 1962.

_____. Prueba de Lectura, Serie Interamericana, Primario, Nivel 1, Forma CEs. Austin, Texas: Guidance Testing Associates, 1966.

_____. Prueba de Lectura, Serie Interamericana, Primario, Nivel 2, Forma CEs and Forma DEs. Austin, Texas: Guidance Testing Associates, 1962.

_____. Test of General Ability, Inter-American Series, Primary, Level 1, Form DE. Austin, Texas: Guidance Testing Associates, 1962.

Manuel, Herschel T. Test of Reading, Inter-American Series, Primary, Level II. Austin, Texas: Guidance Testing Associates, 1962.

_____. Test of Reading, Inter-American Series, Primary, Levels 1 and 2, Form CE and Form DE. Austin, Texas: Guidance Testing Associates, 1962.

Murphy, Helen A. and Durrell, Donald D. Murphy-Durrell Diagnostic Reading Readiness Test, Revised edition. New York: Harcourt, Brace and World, 1964.

Southwest Educational Development Laboratory. Ott Test of Oral Language. Austin, Texas: Southwest Educational Development Laboratory, 1970.

Stemmler, Anne O. The LCT, Language-Cognition Test, Research edition. Austin, Texas: (Mimeographed), 1967.

Thurstone, L.L. and Jeffrey, T.E. Identical Forms, Research edition. Chapel Hill: The Psychometric Laboratory, The University of North Carolina, 1956.

_____. Pattern Copying, Research edition released by Thelma G. Thurstone.

VITA

Mary Ellen Swanson Salinas was born in Iowa City, Iowa, on May 29, 1942, the daughter of Ellen Price Swanson and Leslie William Swanson. She attended the elementary and secondary schools in Mason City, Iowa, graduating from Mason City High School in 1960. She received the degree of Bachelor of Arts with a major in economics in 1964 from Carleton College, Northfield, Minnesota. In September, 1965, she entered the Graduate School of The University of Texas at Austin. She was awarded the degree of Master of Arts in economics in January, 1969. She was employed by the Austin Independent School District as a third grade teacher during the 1969-70 school year. Between July, 1970, and June, 1972, she was employed by the Southwest Educational Development Laboratory in Austin, Texas, as an Evaluation Coordinator and a Staff Development Specialist. In 1970 she married Fernando Enrique Salinas of La Joya, Texas.

Permanent address: 204 North Taylor
Mason City, Iowa 50401

This dissertation was typed by Elizabeth A. Davis.